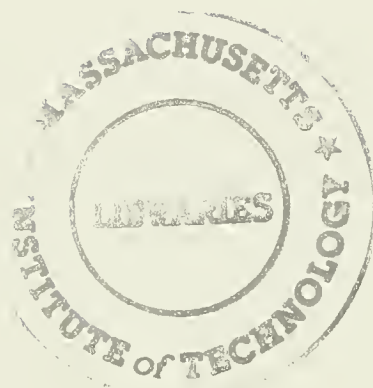






Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

<http://www.archive.org/details/consistencyasyp00newe>

HB31
.M415

no. 584

**working paper
department
of economics**

Consistency and Asymptotic Normality
of
Nonparametric Projection Estimators

Whitney K. Newey

No. 584

Rev. July 1991

**massachusetts
institute of
technology**

**50 memorial drive
cambridge, mass. 02139**



Consistency and Asymptotic Normality
of
Nonparametric Projection Estimators

Whitney K. Newey

No. 584

Rev. July 1991

DEWEY

M.I.T. LIBRARIES

NOV 14 1991

RECEIVED

Consistency and Asymptotic Normality of Nonparametric Projection Estimators

Whitney K. Newey
MIT Department of Economics
March, 1991
Revised, July, 1991

Abstract

Least squares projections are a useful way of describing the relationship between random variables. These include conditional expectations and projections on additive functions. Sample least squares provides a convenient way of estimating such projections. This paper gives convergence rates and asymptotic normality results of least squares estimators of linear functionals of projections. General results are derived, and primitive regularity conditions given for power series and splines. Also, it is shown that mean-square continuity of a linear functional is necessary for $\sqrt{n}$ -consistency and sufficient under conditions for asymptotic normality, and this result is applied to estimating the parameters of a finite dimensional component of a projection and to weighted average derivatives of projections.

Keywords: Nonparametric regression, additive interactive models, partially linear models, average derivatives, polynomials, splines, convergence rates, asymptotic normality.

1. Introduction

Least squares projections of a random variable y on functions of a random vector q provide a useful way of describing the relationship between y and q . The most familiar example is the conditional expectation $E[y|q]$, which is the projection on the linear space of all (measurable, finite mean-square) functions of q . Estimation of this projection is the nonparametric regression problem. Motivated partly by the difficulty of estimating $E[y|q]$ when q has high dimension, projections on smaller sets of functions have been considered, by Breiman and Stone (1978), Breiman and Friedman (1985), Friedman and Stuetzle (1981), Stone (1985), Zeldin and Thomas (1977). These include projections on the set of functions that are additive in linear combinations of q , and generalizations to allow the component functions to be multi-dimensional.

One simple way to estimate nonparametric projections is by regression on a finite dimensional subset, with dimension allowed to grow with the sample size, e.g. as in Agarwal and Studden (1980), Gallant (1981), Stone (1985), Cox (1988), and Andrews (1991), which will be referred to here as series estimation. This type of estimator may not be good at recovering the "fine structure" of the projection relative to other smoothers, e.g. see Buja, Hastie, and Tibshirani (1989), but is computationally simple. Also, the fine structure is less important for mean-square continuous functionals of the projection, such as the parameters of partially linear models or weighted average derivative (examples discussed below), which are essentially averages.

This paper derives convergence rates and asymptotic normality of series estimators of projection functionals. Convergence rates are important because they show how dimension affects the asymptotic accuracy of the estimators (e.g. Stone 1982, 1985) and are useful for the theory of semiparametric

estimators that depend on projection estimates (e.g. Newey 1991). Asymptotic normality is useful for statistical inference about functionals of projection, such as derivatives. The paper gives mean-square rates for estimation of the projection and uniform convergence rates for estimation of functions and derivatives. Conditions for asymptotic normality and consistent estimation of asymptotic standard errors are given, and applied to estimation of a component of an additive projection and its derivatives. Fully primitive regularity conditions are given for power series and spline regression, as well as more general conditions that may apply to other types of series. The regularity conditions allow for dependent observations, so that they are of use for time series models.

The paper also relates continuity properties of linear functionals of projections to $\sqrt{n}$ -consistent estimation. Under a regularity condition on the projection residual variance, continuity in mean-square is shown to be necessary for existence of a (regular) $\sqrt{n}$ -consistent estimator, and sufficient for asymptotically normal series estimators. This result is used to derive $\sqrt{n}$ -consistency results for partially linear models with an additive nonparametric component (a generalization of the model of Engle et. al., 1984) with and for weighted average derivatives of (possibly) additive models (a generalization of the Stoker (1986) functional).

One problem that motivates the results given here is estimation of an additive nonparametric autoregression,

$$(1.1) \quad y_t = h_1(y_{t-1}) + \cdots + h_r(y_{t-r}) + \varepsilon_t,$$

where ε_t is the residual from the projection of y_t on additive functions of r lags. This model avoids high dimensional arguments but allows for several lags, which seems useful for short time series. The convergence rates

and asymptotic normality results apply to estimates of this projection, although asymptotic normality here requires it to be a dynamic regression, where $E[\varepsilon_t | y_{t-1}, y_{t-2}, \dots] = 0$. Effects of lagged values can be quantified by the weighted average derivative $\int w(y_{t-j}) [\partial h(y_{t-j}) / \partial y_{t-j}] dy_{t-j}$, ($j = 1, \dots, r$). The results to follow include primitive regularity conditions for $\sqrt{n}$ -consistency and asymptotic normality of series estimators of this functional, and could also be applied to generalizations that allow interactions between lags in equation (1.1).

These results are an addition to previous work on the subject, including that of Agarwal and Studden (1980), Gallant (1981), Stone (1985), Cox (1988), Andrews (1991), Andrews and Whang (1990), because the results (including asymptotic normality) do not require that the projection equal the conditional expectation, similarly to Stone (1985), but not to the others. In addition, the results allow for dependent observations, and apparently improve in some respects on those of Cox (1988) and Andrews (1991) for the special case of conditional expectations. There is some overlap of the convergence rates with a recent paper by Stone (1990) on additive-interactive spline estimators, which the author saw only after the first version of this paper was written. Stone's rate results (1990) are implied by those of Section 7 below, under conditions that are weaker in some respects (allow for dependence) and stronger in others (imposing a side condition on allowed number of terms). Also, the same convergence rate result is given in Section 6 for variable degree polynomial regression, which is not considered in Stone (1990), and uniform rates and asymptotic normality are shown here.

2. Series Estimators

The results of this paper concern estimators of least squares projections that can be described as follows. Let z denote a data observation, y and q (measurable) functions of z , with q having dimension r . Let $\mathcal{H}$ denote a mean-squared closed, linear subspace of the set of all functions of q with finite mean-square. The projection of y on $\mathcal{H}$ is

$$(2.1) \quad h(q) = \operatorname{argmin}_{\tilde{h} \in \mathcal{H}} E[\{y - \tilde{h}(q)\}^2].$$

An example is the conditional expectation, $h(q) = E[y|q]$, where $\mathcal{H}$ is the set of all measurable functions of q with finite mean-square. An important generalization has $q = (q'_1, q'_2)'$, q_1 has finite mean-square, $\tilde{q}_{2\ell}$, ($\ell = 1, \dots, L$) are subvectors of q_2 , and

$$(2.2) \quad \mathcal{H} = \{q'_1 \tilde{\beta} + \sum_{\ell=1}^L \tilde{h}_{2\ell}(\tilde{q}_{2\ell}) : \tilde{\beta} \in \mathbb{R}^S, E[\tilde{h}_{2\ell}(\tilde{q}_{2\ell})^2] < \infty\}.$$

Primitive conditions for this set to be closed are given in Section 6. This is a smaller set of functions, whose consideration is motivated partly by the difficulty of estimating conditional expectations for x with many dimensions; e.g. see Stone (1985) for discussion and references. The general theory allows for any closed $\mathcal{H}$, (e.g. $\mathcal{H} = \{w(q)[\sum_{\ell} \tilde{h}_{\ell}(\tilde{q}_{\ell})]\}$, $w(q)$ a known function, under conditions for this to be closed), and primitive conditions are given for power series and spline estimators of a projection on the $\mathcal{H}$ of equation (2.2).

The estimators of $h(q)$ considered here are sample projections on a finite dimensional subspace of $\mathcal{H}$, which can be described as follows. Let $p^K(q) = (p_{1K}(q), \dots, p_{KK}(q))'$ be a vector of functions, each of which is an element of $\mathcal{H}$. Denote the data observations by y_i and q_i ,

($i = 1, 2, \dots$), and let $y \equiv (y_1, \dots, y_n)'$ and $p^K \equiv [p^K(q_1), \dots, p^K(q_n)]$, for sample size n . An estimator of $h(q)$ is

$$(2.3) \quad \hat{h}(q) = p^K(q)' \hat{\pi}, \quad \hat{\pi} = (p^K, p^K)^{-K} p^K y,$$

where $(\cdot)^{-}$ denotes a generalized inverse, and K subscripts for $\hat{h}(q)$ and $\hat{\pi}$ have been suppressed for notational convenience. The matrix p^K, p^K will be asymptotically nonsingular under conditions given below, making the choice of generalized inverse asymptotically irrelevant.

Often the object of interest is not the projection, but rather some functional $A(h)$ of $h(q)$, where $A(h)$ is an $s \times 1$ vector of real numbers. Examples are $A(h) = \partial^\lambda h(\bar{q})$, a partial derivative evaluated at some point $\bar{q}$, and $A(h) = h_j(\tilde{q}_j) - h_j(\bar{q}_j) = h(\bar{q}_1, \dots, \tilde{q}_j, \dots, \bar{q}_r) - h(\bar{q})$ for $\mathcal{H} = \{\sum_{j=1}^r \tilde{h}_j(q_j)\}$, the difference of the j th component of an additive projection, at two different points.

Another example is the parameters of the finite dimensional component of a projection, i.e. β in $h(q) = q_1' \beta + \sum_{\ell=1}^L h_{2\ell}(\tilde{q}_{2\ell})$. Estimators of such parameters have been analyzed by Chamberlain (1986), Heckman (1986), Rice (1986), Robinson (1988), Schick (1986), and others, but only under the conditions $h(q) = E[y|q]$ and $L = 1$. Here the nonparametric component can be additive, which leads to a more efficient estimator if $h(q) = E[y|q]$ and $\text{Var}(y|q)$ is constant: see Section 5. Let $\mathcal{H}_2 = \{\sum_{\ell=1}^L \tilde{h}_{2\ell}(\tilde{q}_{2\ell})\}$, assume for the moment that $\mathcal{H}_2$ is closed, let $P(q_1|\mathcal{H}_2)$ denote the vector of projections of each element of q_1 on $\mathcal{H}_2$, and $M = E[\{q_1 - P(q_1|\mathcal{H}_2)\}\{q_1 - P(q_1|\mathcal{H}_2)\}']$. Assuming that M is nonsingular, the identification condition for β , it follows that

$$(2.4) \quad \beta = A(h), \quad A(h) = M^{-1} E[\{q_1 - P(q_1|\mathcal{H}_2)\} h(q)].$$

In addition, once β is in hand, one can consider linear functionals of $h_2(q_2)$, by subtracting off $q_1'\beta$ for any q_1 , e.g. for $L = 1$, $\sum_{\ell=1}^L h_{2\ell}(\bar{q}_{2\ell}) h(q_1, \bar{q}_2) - q_1' M^{-1} E[\{q_1 - P(q_1 | \mathcal{H}_2)\} h(q)]$, where $\bar{q}_2$ is some specified value of q_2 .

Another interesting example is a weighted integral of a partial derivative of $h(q)$, of the form

$$(2.5) \quad A_j(h) = \int w_j(q) \theta^{\lambda_j} h(q) dq, \quad (j = 1, \dots, s),$$

for a multi-indices λ_j and functions $w_j(q)$. This is an average derivative functional similar to that of Stoker (1986), including the nonparametric autoregression from Section 1. Estimators of similar functionals have been analyzed by Hardle and Stoker (1989), Powell, Stock, and Stoker (1989), and Andrews (1991).

In this paper most of the analysis will concern linear functionals of h , such as the above examples. The natural "plug-in" estimators of linear functionals have a simple form. Let $A = (A(p_{1K}(q)), \dots, A(p_{KK}(q)))'$. Because $\hat{h}(q)$ is a linear combination of elements of $\mathcal{H}$, linearity of $A(h)$ implies

$$A(\hat{h}) = A' \hat{\pi}.$$

The paper focuses on linear functionals because the linearity in $\hat{\pi}$ of this estimator leads to straightforward asymptotic distribution theory for $A(\hat{h})$. Of course, the delta-method can also be used to analyze nonlinear functions of such estimators.

The idea of sample projection estimators is that they should approximate $h(q)$ if K is allowed to grow with the sample size. The two key features of this approximation are that 1) each component of $p^K(q)$ is an element of $\mathcal{H}$, and 2) $p^K(q)$ "spans" $\mathcal{H}$ as K grows (i.e. for any function in $\mathcal{H}$, K can

be chosen big enough that there is a linear combination of $p^K(q)$ that approximates it arbitrarily closely in mean square). Under 1), $\hat{\pi}$ estimates $\pi \equiv (E[p^K(q_1)p^K(q_1)'])^{-1}E[p^K(q_1)y_1] = (E[p^K(q_1)p^K(q_1)'])^{-1}E[p^K(q_1)h_0(q_1)]$, the coefficients of the projection of $h_0(q)$ on $p^K(q)$. Thus, under 1) and 2), $p^K(q)'\pi$ will approximate $h_0(q)$. Consequently, when the estimation error in $\hat{\pi}$ is small, $\hat{h}(q)$ should approximate $h_0(q)$.

Two types of approximating functions will be considered in detail. They are:

Power Series: Let $\lambda = (\lambda_1, \dots, \lambda_r)'$ denote an r -dimensional vector of nonnegative integers, i.e. a multi-index, with norm $|\lambda| = \sum_{j=1}^r \lambda_j$, and let $q^\lambda \equiv \prod_{\ell=1}^r q_\ell^{\lambda_\ell}$. For a sequence $(\lambda(k))_{k=1}^\infty$ of distinct such vectors, a power series approximation corresponds to

$$(2.6) \quad p_{kK}(q) = \begin{cases} q_{1k}, & k = 1, \dots, s \\ q_2^{\lambda(k-s)}, & k = s+1, \dots \end{cases},$$

allowing for the finite dimensional component of $\mathcal{H}$ discussed above.

Throughout the paper it will be assumed that $\lambda(k-s)$ are ordered in the natural way, with the degree of the terms $|\lambda(k-s)|$ monotonically increasing in k . Also, an additive q_2 component of the projection will be allowed by including in $\{q_2^{\lambda(k-s)}\}$ only those terms with components that are subvectors of some $\tilde{q}_{2\ell}$, i.e. by requiring that for each multi-index $\lambda(k-s)$ there exists a $\tilde{q}_{2\ell}$ such that the only nonzero components of $\lambda(k-s)$ are those where the corresponding component of q is included in $\tilde{q}_\ell$. The spanning condition will be that all such terms appear in $\{\lambda(k-s)\}$. All of these requirements are summarized in the statement that $\{q_2^{\lambda(k-s)}\}$ consists of all multivariate powers of each $\tilde{q}_{2\ell}$, ordered so that $|\lambda(k-s)|$ is monotonic increasing.

The theory to follow uses orthogonal polynomials, which may also have computational advantages. If each $q_2^{\lambda(k-s)}$ is replaced with the product of orthogonal polynomials of order corresponding to components of $\lambda(k-s)$, with respect to some weight function on the range of q_1 , and the distribution of q_1 is similar to this weight, then there should be little collinearity among the different $q_{2i}^{\lambda(k-s)}$. The estimator will be numerically invariant to such a replacement (because $|\lambda(k-s)|$ is monotonically increasing), but it may alleviate the well known multicollinearity problem for power series.

Splines: Splines, which are smooth piecewise polynomials, can be formulated as projections if their knots (the joining points for the polynomials) are held fixed. They have attractive features relative to power series, being less oscillatory and less sensitive to bad approximation over small regions. The theory here requires that the knots be placed in the support of q_{2i} , which therefore must be known. For convenience the support is normalized to be $\mathcal{J} = \prod_{j=1}^{r-s} [-1, 1]$, and attention restricted to evenly spaced knots. Let $(\cdot)_+ = 1(\cdot > 0)(\cdot)$. An m^{th} degree spline with $L+1$ evenly spaced knots on $[-1, 1]$ is a linear combination of

$$\rho_{kL}(u) = \begin{cases} u^k, & 0 \leq k \leq m, \\ \{[u + 1 - 2(k-m)/(L+1)]_+\}^m, & m+1 \leq k \leq m+L. \end{cases}$$

For a set of multi-indices $\{\lambda(k)\}$, with $\lambda_j(k) \leq m+L-1$ for each j and k , the approximating functions for q_2 will be products of univariate splines, i.e.

$$(2.7) \quad p_{kK}(q) = \begin{cases} q_{1k}, & k = 1, \dots, s \\ \prod_{j=1}^{r-s} \rho_{\lambda_j(k-s), L_j}(q_{2j}), & k = s+1, \dots, K \end{cases},$$

Note that implicit in K is a choice of number of knots for each of the

components of q_2 and a choice of which multiplicative components to include. Throughout the paper it will be assumed that the the ratio of numbers of knots for each pair of elements of q is bounded above and below. An additive q_2 component of the projection will be allowed for by requiring that the multi-indices satisfy the same condition as for power series, which can be summarized in the statement that the q_2 terms consist of all interactions of components that appear in any $\tilde{q}_{2\ell}$.

The condition that the support of q_2 is $\mathcal{J}$ is not as restrictive as it may first appear. Suppose that there are "original" variables x_2 with support $\mathbb{R}^{r-s}$, and $\tau(\cdot)$ is a univariate one-to-one transformation with domain $\mathbb{R}$ and range $[-1,1]$, then $q_2 = (\tau(x_{21}), \dots, \tau(x_{2, r-s}))'$ will have support $\mathcal{J}$. Since additive projections are invariant to such componentwise, one-to-one transformations, the spline estimator based on q_2 will also estimate the original projection. Of course, the condition that the support of x_2 is $\mathbb{R}^{r-s}$ is restrictive. Also, the bounds on derivatives of the projection imposed to obtain convergence rates in what follows are restrictive. The transformation must be continuously differentiable with positive derivative to preserve differentiability of the projection, and boundedness of the derivatives will require that the derivatives of the original projection go to zero as x_2 grows faster than the derivatives of the transformation. For example, if $\tau(\cdot) = 2F(\cdot) - 1$, where $F(x)$ is a CDF that is continuously differentiable of all orders, then the order of differentiability of the projection is preserved under the transformation, but boundedness of derivatives requires that the derivatives of the projection go to zero faster than the density of F as x_2 goes to infinity.

Fixed, evenly spaced knots is restrictive, and is motivated by theoretical convenience. A judicious choice of transformation may help alleviate effects of evenly spaced knots. If a distribution function is used

to transform the data, as discussed above, and the distribution matches closely the true distribution, then the transformed variable will be "spread out," which can improve splines with fixed evenly spaced knots. Allowing for estimated knots (e.g. via smoothing splines, as in Wahba, 1984) is known to lead to more accurate estimates, but is outside the scope of this paper.

The theory to follow uses B-splines, which are a linear transformation of the above basis that is nonsingular on $[-1,1]$ and has low multicollinearity. The low multicollinearity of B-splines and recursive formula for calculation also leads to computational advantages; e.g. see Powell (1981).

Series estimates depend on the choice of the number of terms K , so that it is desirable to choose K based on the data. With a data-based choice of K , these estimates have the flexibility to adjust to conditions in the data. For example, one might choose K by delete one cross validation, by minimizing the sum of squared residuals $\sum_{i=1}^n [y_i - \hat{h}_{-iK}(x_i)]^2$, where $\hat{h}_{-iK}(x_i)$ is the estimate of the regression function computed from all the observations but the i^{th} . Some of the results to follow will allow for data based K .

3. Regularity Conditions

This section lists and discusses some fundamental regularity conditions on which all the following results are based. The first Assumption limits dependence of the observations.

Assumption 3.1: $\{(y_i, q_i)\}$ is stationary and α -mixing with mixing coefficients $\alpha(t) = O(t^{-\mu})$, $(t = 1, 2, \dots)$, for $\mu > 2$.

The stationarity assumption could be relaxed, but is not done so in order to keep the notation as simple as possible. The results will also make use of moment conditions on y . Let $u = y - h_0(q)$, $u_i = y_i - h_0(q_i)$. Also, for a matrix D let $\|D\| = [\text{trace}(D'D)]^{1/2}$, for a random matrix Y , $|Y|_v = \{E[\|Y\|^v]\}^{1/v}$, $v < \infty$, and $|Y|_\infty$ the infimum of constants C such that $\text{Prob}(\|Y\| < C) = 1$.

Assumption 3.2: $|u_i|_s$ is finite for $s \geq 2$ and $E[u_i^2 | q_i]$ is bounded.

The bounded second conditional moment assumption is quite common in the literature (e.g. Stone, 1985). Apparently it can be relaxed only at the expense of affecting the convergence rates (e.g. see Newey, 1990), so to avoid further complication this assumption is retained.

Assumption 3.3: Either a) z_i is uniform mixing with mixing coefficients $\phi(t) = O(t^{-\mu})$, $(t = 1, 2, \dots)$, for $\mu > 2$ or; b) there exists $c(t)$ such that $|E[u_i u_{i+t} | q_i, q_{i+t}]| \leq c(t)$ and $\sum_{t=1}^{\infty} c(t) < \infty$.

This assumption is restrictive, but covers many cases of interest, including a dynamic nonparametric regression with $h_0(q_i) = E[y_i | q_i, y_{i-1}, q_{i-1}, y_{i-2}, \dots]$.

The next Assumption is useful for controlling $(p^{K,p}/n)^{-}$.

Assumption 3.4: For the support Q of q_1 i) For each K there is $\{P_{kK}(q)\}_{k=1}^K$ and a nonsingular, constant matrix A_K with $p^K(q)' = (P_{1K}(q), \dots, P_{KK}(q))A_K$ for all $q \in Q$; ii) There is $\zeta_0(K)$ such that $\max_{k \leq K} |P_{kK}(q_1)| \leq \zeta_0(K)$; iii) For $P^K(q) = (P_{1K}(q), \dots, P_{KK}(q))'$ there is a probability measure $\tilde{P}$ with $P(q_1 \in Q) \geq c\tilde{P}(q_1 \in Q)$ for any measurable set $Q \subseteq Q$ and the smallest eigenvalue of $\int_Q P^K(q)P^K(q)' d\tilde{P}(q)$ is bounded away from zero as $K \rightarrow \infty$.

The bounds in ii) give a convergence rate for $p^{K,p^K/n}$, while iii) controls its singularity. Hypothesis iii) is essentially a normalization, that loads all of the restrictions onto ii). Without this type of normalization the second moment matrix can be ill-conditioned, leading to technical difficulties. For example, if $Q = \mathbb{R}$, q is uniformly distributed on $[0,1]$, and $p_{kK}(q) = q^{k-1}$, then $E[p^K(q)p^K(q)'] = [\sigma_{ij}]$, $\sigma_{ij} = 1/(i+j-1)$, which has a smallest eigenvalue that goes to zero faster than K factorial.

One approach to verifying this assumption is to find a lower bound $\lambda(K)$ on the smallest eigenvalue of $E[p^K(q)p^K(q)']$, and then let $P_{kK}(q) = p_k(q)[\lambda(K)]^{-1/2}$, as in Newey (1988a) for power series. Another approach, is to let $P_{kK}(x)$ be a transformation that is orthonormal with respect to some density and assume that the true distribution dominates the one corresponding to that density, and to use known bounds for orthonormal functions, as in Cox (1988), Newey (1988b), and Andrews (1991) for power series. A third approach to find a transformation that is well conditioned, though not orthogonal, as for B-splines in Agarwal and Studden (1980) and Stone (1985).

The next Assumption specifies the way in which the number of terms is allowed to depend on the data.

Assumption 3.5: $K = \hat{K}(z_1, \dots, z_n, n)$ such that i) There are $\underline{K}(n) \leq \bar{K}(n)$ with $\underline{K}(n) \leq \hat{K} \leq \bar{K}(n)$ with probability approaching one; ii) $p^{\underline{K}(n)}(q)$ is a subvector of $p^K(q)$, which is a subvector of $p^{\bar{K}(n)}(q)$, for all $\underline{K}(n) \leq K \leq \bar{K}(n)$.

That is, $\hat{K}$ is allowed to be random, but must lie between nonrandom upper and lower limits with probability approaching one, and the approximating functions must be nested. Nonrandom K is included as a special case where $\underline{K} = \bar{K}$. These upper and lower limits control variance and bias, respectively (the larger is K the less bias there is but the more variance). It would be interesting to derive such upper and lower limits for specific choices of $\hat{K}$ (e.g. cross-validation), but these results are beyond the scope of this paper.

Some results below will require that the transformed approximating functions are also nested:

Assumption 3.6: Assumption 3.4 is satisfied and for $P^K(q)$ of Assumption 3.3, $p^{\underline{K}(n)}(q)$ is a subvector of $P^K(q)$, which is a subvector of $P^{\bar{K}(n)}(q)$, for all $\underline{K}(n) \leq K \leq \bar{K}(n)$.

This Assumption is satisfied for power series but not for splines, so that the following spline results are limited to the nonrandom K case.

The next Assumption imposes a rate condition for convergence of the second moment matrix. Let $\tilde{K}(n)$ denote the number of elements of $P^{\bar{K}(n)}(q)P^{\bar{K}(n)}(q)'$ that are nonzero for some $q \in Q$, and for notational convenience suppress the n argument in K , $\bar{K}$, $\underline{K}$, and $\tilde{K}$ henceforth.

Assumption 3.7: $\tilde{K}\zeta_0(\bar{K})^4/n \rightarrow 0$.

This is a side condition that will be maintained throughout. It limits the growth rate of K in a way that may be nonoptimal in the mean-square error sense, as discussed in Sections 6 and 7, although it is weaker than similar side conditions imposed by Cox (1988) and Andrews (1991).

The bias of these estimators depends on the error from the finite dimensional approximation. Sobolev norms will be used to quantify this error. For a measurable function $f(q)$ defined on Q , let

$$|f(q)|_{d,v} \equiv \max_{|\lambda| \leq d} |\partial^\lambda f(q_i)|_v = \max_{|\lambda| \leq d} \{E[|\partial^\lambda f(q_i)|^v]\}^{1/v}$$

$$|f(q)|_{d,\infty} \equiv \max_{|\lambda| \leq d} \max_{q \in Q} |\partial^\lambda f(q)|.$$

The norm $|f(q)|_{d,v}$ will be taken to be infinite if $\partial^\lambda f(q)$ does not exist for some $|\lambda| \leq d$. Inclusion of derivatives in these norms will be useful for deriving properties of $\partial^\lambda \hat{h}(q)$.

Many of the results will be based on the following polynomial approximation rate hypothesis:

Assumption 3.8: For each class of functions $\mathcal{F}$, there exists $C = C(\mathcal{F}, d, v)$ and $\alpha = \alpha(\mathcal{F}, d, v)$ such that for all $f \in \mathcal{F}$,

$$\inf_{\pi \in \mathcal{R}} |f(q) - p^K(q)' \pi|_{d,v} \leq CK^{-\alpha}.$$

This condition is not primitive, but is known to be satisfied in many cases. Primitive conditions for power series and splines are given in Sections 6 and 7.

In order for the same bias bounds to apply to estimated functionals of interest, it is necessary that $A(h)$ be continuous with respect to the same norm as in Assumption 3.8, which is imposed in the following condition.

Assumption 3.9: For the d and v of Assumption 3.8, $A(h)$ is a continuous linear functional with respect to the Sobolev norm $|h|_{d,v}$, i.e. there is C such that for all $h \in \mathcal{H}$, $\|A(h)\| \leq C|h|_{d,v}$.

4. Convergence Rates

This Section gives mean-square convergence rates for $\hat{h}(q)$ and uniform consistency rates for its derivatives and continuous linear functionals. The results include both sample and population mean-square error rates.

Theorem 4.1: If and Assumptions 3.1 - 3.5 and 3.7 are satisfied for $d = 0$, and $\mathcal{F} = \{h\}$, then

$$\sum_{i=1}^n [\hat{h}(q_i) - h(q_i)]^2 / n = O_p(\bar{K}/n + \underline{K}^{-2\alpha} \{ \sum_{\underline{K} \leq K \leq \bar{K}} (K/\underline{K})^{-\alpha v} \}^{2/v})$$

If Assumption 3.6 is also satisfied, then for the CDF $F(q)$ of q_i ,

$$\int [\hat{h}(q) - h(q)]^2 dF(q) = O_p(\bar{K}/n + \underline{K}^{-2\alpha} \{ \sum_{\underline{K} \leq K \leq \bar{K}} (K/\underline{K})^{-\alpha v} \}^{2/v}).$$

The two terms in the convergence rate essentially correspond to variance and bias. The bias term, which is $\underline{K}^{-2\alpha} \{ \sum_{\underline{K} \leq K \leq \bar{K}} (K/\underline{K})^{-\alpha v} \}^{2/v}$, will be equal to $\underline{K}^{-2\alpha}$ if $\underline{K} = \bar{K}$, and for $v > 1/\alpha$ is bounded above by $\underline{K}^{-2\alpha+2/v}$. A consequence of the second conclusion is convergence rates for some version of the additive components, because $\mathcal{H}$ closed implies that the mapping from $h(q)$ to some decomposition into additive components is mean-square continuous (see Bickel et. al., 1990, Appendix).

Uniform convergence rates depend on bounds for the derivatives of the

series terms.

Assumption 4.1: For each $k \leq K$, $P_{kK}(q)$ is differentiable of order ρ and for all multi-indices λ with $|\lambda| \leq \rho$ there is $\zeta_{|\lambda|}(K)$ such that with probability one $\max_{k \leq K} |\partial^\lambda P_{kK}(q_i)| \leq \zeta_{|\lambda|}(K)$ and $\zeta_{|\lambda|}(K) \geq \zeta_0(K)$,

Theorem 4.2: If Assumptions 3.1 - 3.8 are satisfied, for $d = |\lambda| \leq \rho$, $v = \infty$, and $\mathcal{F} = \{h\}$, and Assumption 4.1 is satisfied, then

$$\sup_{q \in Q} |\partial^\lambda \hat{h}(q) - \partial^\lambda h(q)| = O_p(\bar{K}^{1/2} \zeta_{|\lambda|}(\bar{K}) [(\bar{K}/n)^{1/2} + \underline{K}^{-\alpha}]).$$

The uniform convergence rate for $\hat{h}(q)$ is slower than the mean-square rate and does not attain Stone's (1982) bounds, although it is faster than previously derived rates for series estimators, as further discussed below.

Convergence rates for continuous linear functionals of $\hat{h}(q)$ will follow from this result.

Theorem 4.3: If Assumptions 3.1 - 3.9 and 4.1 are satisfied for $v = \infty$ and $\zeta_{|\lambda|}(K)$ is monotonically increasing in $|\lambda|$, then $A(\hat{h}) - A(h_0) = O_p(\bar{K}^{1/2} \zeta_d(\bar{K}) [(\bar{K}/n)^{1/2} + \underline{K}^{-\alpha}])$.

The implied convergence rate for mean-square continuous linear functions is not sharp, as they are shown to be $\sqrt{n}$ -consistent (under slightly stronger conditions) in Section 5.

5. Asymptotic Normality

An estimator of the asymptotic variance of $A\hat{\pi}$ can be formed from the usual estimator of the asymptotic variance of the projection coefficients $\hat{\pi}$. The asymptotic normality result below will require that the products of elements of $\mathcal{H}$ and the residual be martingale differences, so that no autocorrelation correction is required. Let

$$\hat{\Sigma} \equiv p^K p^K / n, \quad \hat{V} = \sum_{i=1}^n p^K(q_i) p^K(q_i)' [y_i - \hat{h}(q_i)]^2 / n.$$

Treating K as fixed, the White (1980) estimator of the asymptotic variance of the projection coefficient estimator $\hat{\pi}$ is $\hat{\Sigma}^{-1} \hat{V} \hat{\Sigma}^{-1}$. Since $A\hat{\pi}$ is a linear combination of $\hat{\pi}$, a corresponding estimator of the asymptotic variance $A\hat{\pi}$ is

$$\hat{\Omega} = A' \hat{\Sigma}^{-1} \hat{V} \hat{\Sigma}^{-1} A / n.$$

This estimator is consistent as K grows, under conditions to follow.

Further conditions are useful for asymptotic normality of $A(\hat{h})$ and consistency of $\hat{\Omega}$. Let

$$\Sigma = E[p^K(q_i) p^K(q_i)'], \quad V = E[p^K(q_i) p^K(q_i)' u_i^2], \quad \Omega = A' \Sigma^{-1} V \Sigma^{-1} A / n.$$

Assumption 5.1: Assumptions 3.1 - 3.2 are satisfied, with $s > 4\mu/(\mu-1)$, and i) $E[u_i^2 | q_i]$ is bounded away from zero; ii) For any $\tilde{h}(q) \in \mathcal{H}$, $E[\tilde{h}(q_i) u_i | z_{i-1}, z_{i-2}, \dots] = 0$; iii) $K = K(n)$ is nonrandom.

Part ii) of is the martingale difference assumption: it holds if the observations are independent or $h_0(q_i) = E[y_i | z_{i-1}, z_{i-2}, \dots]$.

Assumption 5.2: A has full column rank for some K and for all K there is a nonsingular matrix C_K such that for all $j \leq K$, $C_K P^K(q)$ does not depend on K .

Part ii) of this hypothesis rules out asymptotic linear dependence among different components of $A(\hat{h})$. When $A(h)$ is mean-square continuous in h , a primitive condition for this hypothesis is that $A(h)$ is onto $\mathbb{R}^S$, as discussed below.

Assumption 5.3: Assumption 3.4, 3.8, and 3.9 are satisfied for $v \geq 2$, $\tilde{K} \geq K$, $K\tilde{K}\zeta_0(K)^4/n \rightarrow 0$, and $\sqrt{n}K^{-\alpha} \rightarrow 0$.

The second condition requires essentially that the bias converges to zero faster than $1/\sqrt{n}$ (see Assumption 3.8), which is stronger than the natural condition that the bias go to zero faster than the variance.

Theorem 5.1: If Assumptions 5.1 - 5.3 are satisfied then

$$\Omega^{-1/2}[A(\hat{h}) - A(h)] \xrightarrow{d} N(0, I), \quad \Omega^{-1/2}[\hat{\Omega} - \Omega]\Omega^{-1/2} \xrightarrow{p} 0.$$

Furthermore, if there exists a scalar $\psi_n > 0$ and nonsingular Ω_0 such that $\psi_n \Omega \rightarrow \Omega_0$, then

$$\psi_n^{1/2}[A(\hat{h}) - A(h)] \xrightarrow{d} N(0, \Omega_0), \quad \psi_n \hat{\Omega} \xrightarrow{p} \Omega_0.$$

This result improves on Andrews (1991) in applying to estimators of projections other than conditional expectations, allowing for dependence, and having a faster growth rate for K , but restricts K to be fixed: It is more difficult to allow for random K with dependent observations.

The second conclusion is useful for forming asymptotic Gaussian

confidence intervals in the usual way. If the hypotheses of Theorem 4.3 are satisfied and $\bar{K}^{1/2} \zeta_d(\bar{K})[(\bar{K}/n)^{1/2} + \underline{K}^{-\alpha}] \rightarrow 0$, so that $A(\hat{h})$ is a consistent estimator of $A(h)$, the delta-method can be used to make inference about smooth nonlinear functions of $A(h)$ in the standard way. The hypothesis about ψ_n is not restrictive when $A(h)$ is a scalar, where $\psi_n = \Omega^{1/2}$ will satisfy the hypotheses. However, when $A(h)$ is a vector, it requires essentially that the variance of each component of $A(\hat{h})$ converge to zero at the same rate, which may not be true when, e.g., $A(h)$ includes both $h(q)$ and its derivatives at a single point. It is possible to derive a primitive condition for $\psi_n = n$, corresponding to $\sqrt{n}$ -consistency of $A(\hat{h})$, which is stated in the following result.

Theorem 5.2: Suppose that i) Assumptions 5.1 and 5.3 are satisfied for $d = 0$ and $v = 2$; ii) for any $\tilde{h}(q) \in \mathcal{H}$ there exists π_K such that $E[\{\tilde{h}(q) - p^K(q)' \pi_K\}^2] \rightarrow 0$ as $K \rightarrow \infty$; iii) there exists an $s \times 1$ vector $\delta(q)$ of elements of $\mathcal{H}$ such that $E[\delta(q)\delta(q)']$ exists and is nonsingular, and $A(\tilde{h}) = E[\delta(q)\tilde{h}(q)]$ for all $\tilde{h} \in \mathcal{H}$. Then for $\Omega_0 = E[\sigma^2(q)\delta(q)\delta(q)']$

$$\sqrt{n}[A(\hat{h}) - A(h)] \xrightarrow{d} N(0, \Omega_0), \quad n\hat{\Omega} \xrightarrow{p} \Omega_0.$$

Hypothesis ii) is the minimal mean-square spanning requirement for consistency of the sample projection. By the multivariate Riesz representation theorem (e.g. see Hansen, 1985), hypothesis iii) is equivalent to the statement that $A(h)$ is mean-square continuous and has range $\mathbb{R}^s$.

Furthermore, mean-square continuity of such linear functionals is a *necessary condition* for a finite semiparametric variance bound for $A(h)$, as in Stein (1956), and hence for existence of a (regular) $\sqrt{n}$ -consistent estimator (Chamberlain, 1986), so that mean-square continuity of $A(h)$ characterizes the $\sqrt{n}$ -consistent case.

Theorem 5.3: If $A(h)$ is a scalar functional that is not mean-square continuous then the semiparametric variance bound for $A(h)$ is infinite.

Theorem 5.2 can be specialized to many interesting examples, including the parameters of the finite dimensional component of the projection and average derivatives. As noted in equation (2.4), the parameters β of $h(q) = q_1'\beta + h_2(q_2)$ satisfy $\beta = E[\delta(q)h(q)]$ for $\delta(q) = M^{-1}[q_1 - P(q_1|\mathcal{H}_2)]$. Furthermore, the mean-square spanning hypothesis of ii) will be satisfied as long as $(p_{s+1,K}(q_2), \dots, p_{KK}(q_2))'$ span $\mathcal{H}_2$, giving the following result:

Theorem 5.4: Suppose that i) Assumptions 5.1 and 5.3 are satisfied; ii) for any $\tilde{h}_2(q_2) \in \mathcal{H}_2$ there exists $\{\pi_{kK}\}_{K \geq k \geq s}$ such that $E[|\tilde{h}_2(q_2) - \sum_{k=s+1, K}^K \pi_{kK} p_{kK}(q_2)|^2] \rightarrow 0$, and iii) $M = E[\{q_1 - P(q_1|\mathcal{H}_2)\}\{q_1 - P(q_1|\mathcal{H}_2)\}']$ is nonsingular. Then for $\Omega_0 = M^{-1}E[\sigma^2(q)\{q_1 - P(q_1|\mathcal{H}_2)\}\{q_1 - P(q_1|\mathcal{H}_2)\}]M^{-1}$,

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Omega_0), \quad n\hat{\Omega} \xrightarrow{p} \Omega_0.$$

Sample projection estimators of β have been previously analyzed by Chamberlain (1986), Andrews (1991), and Newey (1990), but only under $q_1'\beta_0 + h_2(q_2) = E[y|q]$, an unrestricted functional form for $h_2(q_2)$ (e.g. $h_2(q_2)$ could not be additive), and independent observations. One implication of this result is that if $E[y|q] = q_1'\beta_0 + h_2(q_2)$ and $\sigma^2(q) = \text{Var}(y|q)$ is constant, then an estimator that imposes additivity $h_2(q_2)$ will be asymptotically more efficient than one that does not: The asymptotic matrices are $\sigma^2(E[\{q_1 - P(q_1|\mathcal{H}_2)\}\{q_1 - P(q_1|\mathcal{H}_2)\}'])^{-1}$ and $\sigma^2(E[\{q_1 - E[q_1|q_2]\}\{q_1 - E[q_1|q_2]\}'])^{-1}$ respectively, which have a positive semi-definite difference. Thus, although imposing additivity does not improve the convergence rate of $\hat{\beta}$, it can lower its asymptotic variance.

Theorem 5.2 can also be specialized to the average derivative functional of equation (2.4) for certain weights. If there are no boundary terms then integration by parts gives

$$\begin{aligned} A_j(h) &= \int_Q w_j(q) \partial^{\lambda_j} h(q) dq = (-1)^{|\lambda_j|} \int_Q \partial^{\lambda_j} w_j(q) h(q) dq \\ &= E[\delta_j(q) h(q)], \quad \delta_j(q) = (-1)^{|\lambda_j|} P(\partial^{\lambda_j} w_j(q) / f(q) | \mathcal{H})(q), \end{aligned}$$

where $f(q)$ is the density of q . Here, $E[\delta_j(q)^2]$ will be finite if $f(q)$ is not too small relative to $\partial^{\lambda_j} w_j(q)$. Let $\delta(q) = (\delta_1(q), \dots, \delta_s(q))'$. Following Stoker (1986), the previous integration by parts will be valid, and hence $A(h)$ mean-square continuous, under the hypotheses of the following result.

Theorem 5.5: Suppose that i) Assumptions 5.1 and 5.3 are satisfied; ii) for any $\tilde{h}(q) \in \mathcal{H}$ there exists π_K such that $E[\{\tilde{h}(q) - p^K(q)' \pi_K\}^2] \rightarrow 0$ as $K \rightarrow \infty$; iii) Q is convex with nonempty interior, $w_j(q)$ is continuously differentiable to order $|\lambda_j|$ on Q and $\partial^{\tilde{\lambda}} w_j(q) = 0$ on the boundary of Q for all $|\tilde{\lambda}| \leq |\lambda_j|$, $E[\delta(q) \delta(q)']$ exists and is nonsingular. Then for $\Omega_0 = E[\sigma^2(q) \delta(q) \delta(q)']$,

$$\sqrt{n}(A(\hat{h}) - A(h_0)) \xrightarrow{d} N(0, \Omega_0), \quad n\hat{\Omega} \xrightarrow{p} \Omega_0.$$

Primitive conditions for these results for power series and splines are given in the following Sections.

6. Power Series

This Section gives primitive conditions for consistency and asymptotic normality of projections that use power series. Throughout both this and the next Sections, it will be assumed that $\mathcal{H}$ is as specified in equation (2.2), and the following hypothesis is satisfied:

Assumption 6.0: i) For each $\tilde{q}_{2\ell}$, ($\ell = 1, \dots, L$), if $\tilde{q}_2$ is a subvector of ℓ then $\tilde{q}_2 = \tilde{q}_{2\ell'}$ for some ℓ' ; ii) There exists a constant $c > 1$ such that for each ℓ , with the partitioning $q_2 = (\tilde{q}_{2\ell}', \tilde{q}_{2\ell}^C)'$, for any $a(q) > 0$, $c \int a(q) d[F(\tilde{q}_{2\ell}) \cdot F(\tilde{q}_{2\ell}^C)] \geq E[a(q)] \geq c^{-1} \int a(q) d[F(\tilde{q}_{2\ell}) \cdot F(\tilde{q}_{2\ell}^C)]$; iii) Either $\beta = 0$ (i.e. q_1 is not present) or q_1 is bounded and for the closure $\mathcal{H}_2$ of $\{\sum_{\ell=1}^L \tilde{h}_{2\ell}(\tilde{q}_{2\ell})\}$, $E\{q_1 - P(q_1 | \mathcal{H}_2)\{q_1 - P(q_1 | \mathcal{H}_2)\}'\}$ is nonsingular.

Conditions i) and ii) are sufficient for $\{\sum_{\ell=1}^L \tilde{h}_{2\ell}(\tilde{q}_{2\ell})\}$ to be closed.

Boundedness of q_1 can be relaxed, but for brevity is not here.

Power series estimators will be mean-square consistent if the regressor distribution has an absolutely continuous component and the $\bar{K}$ grows slowly enough; see Newey (1988a). To obtain convergence rates it is useful to bound the regressor density below, as follows:

Assumption 6.1: There are finite $q_j^1 > q_j^2$, $\nu_j \geq 0$, ($j = 1, \dots, r$) such that the support of q_{2i} is $\prod_{j=1}^r [q_j^1, q_j^2]$ and the distribution of q_{2i} has absolutely continuous component with density bounded below by $C \prod_{j=1}^r [(q_j^2 - q_j)(q_j - q_j^1)]^{\nu_j}$ on the support.

It is also possible to allow for a discrete regressor with finite support, by including all dummy variables for all points of support of the regressor, and all interactions. Because such a regressor is essentially parametric, and allowing for it does not change any of the convergence rate results, this

generalization will not be considered here.

To state further conditions, let $\{\lambda(k)\}_{k=1}^{\infty}$ denote the sequence of multi-indices used in defining the power series or spline interactions in equations (2.3) and (2.4) respectively, and for any (other) multi-index λ ,

$$\mathcal{J}(k) = \{j : \lambda_j(k) \neq 0\}, \quad J(k) = \#\mathcal{J}(k) \leq r,$$

$$\bar{\nu} = \max_k \sum_{j \in \mathcal{J}(k)} \nu_j / J(k), \quad \bar{\lambda} = \max_k \sum_{j \in \mathcal{J}(k)} \lambda_j / J(k).$$

For power series, Assumption 3.7 follows from

Assumption 6.2: $\bar{K}^{4+4\bar{\nu}}/n \rightarrow 0$.

For $r = 1$ and $\bar{\nu} = 0$, this condition is $\bar{K} = o(n^{1/4})$, which is weaker than Cox's (1988, p. 715) $\bar{K} = o(n^{1/5})$ requirement.

Primitive approximation rate conditions (as in Assumption 3.8) for power series follow from known results of Lorentz (1986), Powell (1981), or a Taylor expansion.

Assumption 6.3: Each of the components $h_{0\ell}(\tilde{q}_\ell)$, ($\ell = 1, \dots, L$), is continuously differentiable of order h on the support of q_1 .

This hypothesis implies Assumption 3.8 for $v = \infty$, with $d = 0$ and $\alpha = h/r$ and with $\alpha = h-d$ when $r = 1$. A literature search has not yet revealed corresponding conditions for $d > 0$ and $r > 1$, but rates for this case follow from a Taylor expansion under the following (strong) condition:

Assumption 6.4: There is a constant C such that for each multi-index λ , the λ^{th} partial derivative of each additive component of $h(q)$ exists and is bounded by $C^{|\lambda|}$.

The first power series result gives convergence rates.

Theorem 6.1: Suppose that Assumptions 3.1 - 3.3 and 6.0 - 6.3 are satisfied.

Then

$$\sum_{i=1}^n [\hat{h}(q_i) - h(q_i)]^2 / n = O_p(\bar{K}/n + \underline{K}^{-2h/\nu}),$$

$$\int [\hat{h}(q) - h(q)]^2 dF(q) = O_p(\bar{K}/n + \underline{K}^{-2h/\nu})$$

$$\sup_{q \in Q} |\hat{h}(q) - h(q)| = O_p(\bar{K}^{1+\bar{\nu}} \{[\bar{K}/n]^{1/2} + \underline{K}^{-h/\nu}\})$$

Suppose, in addition, that either a) $\nu = 1$, $\bar{\lambda} \leq h$, and $\alpha = h - \bar{\lambda}$, or; b)

Assumption 6.4 satisfied and α is any positive number; Then

$$\sup_{q \in Q} |\partial^\lambda \hat{h}(q) - \partial^\lambda h(q)| = O_p(\bar{K}^{1+\bar{\nu}+2\bar{\lambda}} \{[\bar{K}/n]^{1/2} + \underline{K}^{-\alpha}\}).$$

This result implies optimal convergence rates for power series estimators of $h(q)$ when K goes to infinity at the optimal rate and Assumption 6.2 is satisfied. If the density of q is bounded away from zero, $\underline{K} = cn^\gamma$, $\bar{K} = Cn^\gamma$, $\gamma = \nu/(2h+\nu)$, and $h > 3\nu/2$, the mean-square convergence rate for $h(q)$ is $n^{-h/(2h+\nu)}$, which attains Stone's (1982) bounds. The side condition that $h > 3\nu/2$, which is needed to guarantee Assumption 6.7, limits this optimality result, but is weaker than the corresponding condition in Cox (1988). These mean-square error results apply to additive projections (rather than conditional expectations), like Stone (1985, 1990) but unlike Cox (1988) or Andrews and Whang (1990), allows for interactive terms, similarly to Stone (1990), (although Stone (1985) also derives optimal rates for derivatives), and allows for dependent observations. The side condition Assumption 6.2 is not present in Stone (1985) or Andrews and Whang (1990), but it implies a population mean-square error result, unlike Andrews and Whang (1990). In comparison with Cox's (1988) uniform convergence rate of $K^{5/2}/\sqrt{n} + K^{2-h} =$

$K^2([K/n]^{1/2} + K^{-h})$ for univariate q with density bounded away from zero (in Cox's notation, $h = 1$ and $k = 2$), the rate here is the faster $K([K/n]^{1/2} + K^{-h})$, and uniform convergence rates for derivatives are given here.

To state the asymptotic normality result for power series, let $\hat{\Omega}$ denote the variance matrix estimator described in Section 5, for power series.

Theorem 6.2: Suppose that i) Assumptions 3.1, 3.2, 5.1 - 5.2, 6.0, 6.1, are satisfied with $s > 4\mu/(\mu-1)$, $K^{5+4\bar{\nu}}/n \rightarrow 0$; ii) $A(h)$ is continuous with respect to the Sobolev norm $|h|_{d,\omega}$, either a) 6.3 is satisfied, $d = 0$, and $\sqrt{n}K^{-h/\nu} \rightarrow 0$; or b) Assumption 6.3 is satisfied, $\nu = 1$, $\sqrt{n}K^{-h+d} \rightarrow 0$; or c) Assumption 6.4 is satisfied and $K/n^\gamma \rightarrow \infty$ for some $\gamma > 0$; iii) there exists a scalar $\psi_n > 0$ and nonsingular Ω_0 such that $\psi_n \Omega \rightarrow \Omega_0$. Then

$$\psi_n^{1/2} [A(\hat{h}) - A(h)] \xrightarrow{d} N(0, \Omega_0), \quad \psi_n \hat{\Omega} \xrightarrow{p} \Omega_0.$$

In comparison with Andrews (1991), this result applies to projections other than the conditional expectation, or allows for dependent observations, and has weaker growth rate conditions for K : if $\bar{\nu} = 0$ then i) requires $K = o(n^{1/5})$ while Andrews (1991) requires $K = o(n^{1/6})$. For $d = 0$ i) and ii) imply that $h > 5\nu/2$ (e.g. $h(q)$ is thrice continuously differentiable when $\nu = 1$).

This result can be applied to estimation of the components of an additive projection and their derivatives, when the observations are independent.

Theorem 6.3: Suppose the observations are independent, $\sigma^2(q)$ is bounded and bounded away from zero, $|u|_s < \infty$ for $s > 2$, Assumptions 6.0 and 6.1 are satisfied for $v_j = 0$, each $h_j(q_j)$ is continuously differentiable to order h on Q , $K = o(n^{1/5})$, and $\sqrt{n}K^{-h} \rightarrow 0$. Then for any pair of points $\tilde{q}_j, \bar{q}_j$ in the support of q_j , ($j = 1, \dots, r$),

$$\hat{\Omega}^{-1/2}[\{\hat{h}_j(\tilde{q}_j) - \hat{h}_j(\bar{q}_j)\} - \{h_{j0}(\tilde{q}_j) - h_{j0}(\bar{q}_j)\}] \xrightarrow{d} N(0, 1).$$

Also, if $\sqrt{n}K^{-h+d} \rightarrow 0$, then for any $\tilde{q}_j$ in the support of q_j ,

$$\hat{\Omega}^{-1/2}[\partial^d \hat{h}_j(\tilde{q}_j) / \partial q_j^d - \partial^d h_{j0}(\tilde{q}_j) / \partial q_j^d] \xrightarrow{d} N(0, 1).$$

The differencing normalization here is different than the mean centering in Stone (1985), which would be more difficult to work with.

A $\sqrt{n}$ -consistency and asymptotic normality result for power series estimates of mean-square continuous linear functionals is:

Theorem 6.4: Suppose that Assumptions 5.1, 6.0, 6.1, are satisfied and $K^{5+4\bar{\nu}}/n \rightarrow 0$; ii) Assumption 6.3 is satisfied, and $\sqrt{n}K^{-h/\bar{\nu}} \rightarrow 0$; iii) there exists an $s \times 1$ vector $\delta(q)$ of elements of $\mathcal{H}$ such that $E[\delta(q)\delta(q)']$ exists and is nonsingular, and $A(\tilde{h}) = E[\delta(q)\tilde{h}(q)]$ for all $\tilde{h} \in \mathcal{H}$. Then for $\Omega_0 = E[\sigma^2(q)\delta(q)\delta(q)']$,

$$\sqrt{n}[A(\hat{h}) - A(h)] \xrightarrow{d} N(0, \Omega_0), \quad n\hat{\Omega} \xrightarrow{p} \Omega_0.$$

This result can be specialized to the parameters of a finite dimensional component and average derivatives, as follows.

Theorem 6.5: If hypotheses i) and ii) of Theorem 6.5 are satisfied then for

$$\Omega_0 = M^{-1} E[\sigma^2(q)(q_1 - P(q_1 | \mathcal{H}_2))\{(q_1 - P(q_1 | \mathcal{H}_2))'\}] M^{-1},$$

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Omega_0), \quad n\hat{\Omega} \xrightarrow{P} \Omega_0.$$

This result gives fully primitive regularity conditions for $\sqrt{n}$ -consistency and asymptotic normality of a power series estimator of the parameters of a finite dimensional component of a projection. It allows for $h_2(q_2)$ to have the additive form discussed above, and also allows for dependent observations. An analogous result can be given for weighted average derivatives, although for brevity such a result is only given below for splines.

7. Splines

Results for splines are limited to the case

Assumption 7.1: Assumptions 6.0 and 6.1 are satisfied with $\nu_j = 0$, ($j = 1, \dots, J$).

Splines allow for a faster growth rate for the number of terms.

Assumption 7.2: $K^3/n \rightarrow 0$.

Approximation rate conditions for $r = 1$ or $d = 0$ follow from known results, but a literature search has not yet revealed conditions for other cases, which limits the following results.

Theorem 7.1: Suppose that $r = 1$ or $d = 0$, Assumptions 3.1 - 3.3, 7.1, 7.2, and 6.3 are satisfied, and $m \geq h - 1$. Then for $|\lambda| \leq m$,

$$\sum_{i=1}^n [\hat{h}(q_i) - h(q_i)]^2 / n = o_p(K/n + K^{-2h/r}),$$

$$\int [\hat{h}(q) - h(q)]^2 dF(q) = o_p(K/n + K^{-2h/r})$$

$$\sup_{q \in Q} |\hat{h}(q) - h(q)| = o_p(K\{[K/n]^{1/2} + K^{-h/r}\}).$$

If, in addition $r = 1$, then

$$\sup_{q \in Q} |\partial^\lambda \hat{h}(q) - \partial^\lambda h(q)| = o_p(K^{1+\bar{\lambda}}\{[K/n]^{1/2} + K^{-h+\bar{\lambda}}\}).$$

This result yields optimal mean-square convergence rates for spline regression estimation of an additive projection with dependent observations, if $K = n^{1/(2h+1)}$: here the side condition of Assumption 7.2 is satisfied if $h > 1$.

Throughout the rest of Section 7, let $\hat{\Omega}$ be the variance estimator computed as described in Section 5, using splines.

Theorem 7.2: Suppose that $r = 1$ or $d = 0$; i) Assumptions 5.1 - 5.2, 7.1, are satisfied and $K^4/n \rightarrow 0$; ii) $A(h)$ is continuous with respect to the Sobolev norm $|h|_{d,\infty}$ for $d \leq m$, Assumption 6.3 is satisfied for $m \geq h-1$, and $\sqrt{n}K^{-h} \rightarrow 0$; iii) there exists a scalar $\psi_n > 0$ and nonsingular Ω_0 such that $\psi_n \Omega \rightarrow \Omega_0$. Then

$$\psi_n^{1/2} [A(\hat{h}) - A(h)] \xrightarrow{d} N(0, \Omega_0), \quad \psi_n \hat{\Omega} \xrightarrow{p} \Omega_0.$$

Apparently, there are no other asymptotic normality results for spline projections in the literature. The growth rate $K = o(n^{1/4})$ is smaller than $K = o(n^{1/5})$ for power series, so asymptotic normality will require only twice differentiability of $h(q)$ rather than the thrice differentiability for power

series. This result can be specialized analogously to Theorem 6.3, although for brevity this specialization is omitted.

A $\sqrt{n}$ -consistency and asymptotic normality result for spline estimators of mean-square continuous linear functionals is:

Theorem 7.3: Suppose that Assumptions 5.1, 6.1, are satisfied and $K^4/n \rightarrow 0$; ii) Assumption 6.3 is satisfied, $m \geq h-1$ and $\sqrt{n}K^{-h} \rightarrow 0$; iii) there exists an $s \times 1$ vector $\delta(q)$ of elements of $\mathcal{H}$ such that $E[\delta(q)\delta(q)']$ exists and is nonsingular, and $A(\tilde{h}) = E[\delta(q)\tilde{h}(q)]$ for all $\tilde{h} \in \mathcal{H}$. Then for $\Omega_0 = E[\sigma^2(q)\delta(q)\delta(q)']$

$$\sqrt{n}[A(\hat{h}) - A(h)] \xrightarrow{d} N(0, \Omega_0), \quad n\hat{\Omega} \xrightarrow{p} \Omega_0.$$

This result can be specialized to the parameters of a finite dimensional component and average derivatives, as follows. For brevity, only the average derivative result is given here. Let q_{-j} denote the vector of all the components of q other than the j th, and $f(q_j|q_{-j})$ the conditional density of the j th component given the others.

Theorem 7.4: Suppose that hypotheses i) and ii) of Theorem 7.3 are satisfied, for some integer $d > 0$, $d \leq m$, $w_j(q)$ is continuously differentiable to order $|\lambda_j|$ on Q and $\partial^{\tilde{\lambda}} w_j(q) = 0$ on the boundary of Q for all $|\tilde{\lambda}| \leq |\lambda_j|$, $h_{j0}(q_j)$ is continuously differentiable to order d , and $\|[\partial^d w(q_j)/\partial q_j^d]/f(q_j|q_{-j})\|_2 < \infty$. Then

$$\begin{aligned} & \sqrt{n} \left[\int_{[-1,1]} w(q_j) [\partial^d \hat{h}_j(q_j)/\partial q_j^d] dq_j - \int_{[-1,1]} w(q_j) [\partial^d h_{j0}(q_j)/\partial q_j^d] dq_j \right] \\ & \xrightarrow{d} N(0, E[\sigma(q)^2 \delta(q)^2]), \quad n\hat{\Omega} \xrightarrow{p} E[\sigma(q)^2 \delta(q)^2]. \end{aligned}$$

8. Proofs of Theorems

The proofs of Sections 8 and 9 are abbreviated, with details provided only for central or potentially unfamiliar results. A longer version of these sections is available from the author upon request. Throughout, let C be a generic positive constant and $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ be minimum and maximum eigenvalues of a symmetric matrix B . The following Lemmas are useful in proving the results for power series and splines.

Lemma 8.0: If Assumption 6.1 i) and ii) are satisfied, then $\{\sum_{\ell=1}^L \tilde{h}_{2\ell}(\tilde{q}_{2\ell}) : E[\tilde{h}_{2\ell}(\tilde{q}_{2\ell})^2] < \infty, \ell = 1, \dots, L\}$ is closed in mean-square. If in addition Assumption 6.1 iii) is satisfied then $\mathcal{H}$ is closed.

Proof: For now, let $\mathcal{H}_2 = \{\sum_{\ell=1}^L \tilde{h}_{2\ell}(\tilde{q}_{2\ell})\}$, and for notational convenience drop the 2 subscript. By Proposition 2 of Section 4 of the Appendix of Bickel, Klaassen, Ritov, and Wellner (1990), $\mathcal{H}$ closed is equivalent to existence of a constant C such that for each $h \in \mathcal{H}$ there is $h_{\ell}(\tilde{q}_{\ell})$ with $\|h\|_2 \geq C \max_{\ell} \{ \|h_{\ell}\|_2 \}$. Existence of such a C can be shown using an induction argument like that of Stone (1990, Lemma 1, " L_2 Rate of Convergence for Interaction Spline Regression," Tech. Rep. No. 268, Berkeley). As noted there, assuming that this property holds for each maximal dimension less than r , for each $h \in \mathcal{H}$, there is a unique decomposition $h = \sum_{\ell=1}^L h_{\ell}(\tilde{q}_{\ell})$, such that for all $\tilde{q}_k$ that are strict subvectors of $\tilde{q}_{\ell}$, $E[h_{\ell}(\tilde{q}_{\ell})\delta(\tilde{q}_k)] = 0$ for all measurable functions of $\tilde{q}_k$ with finite mean-square. Then, following Stone (1990), it suffices to show that for any "maximal" $\tilde{q}_k$, that is not a proper subvector of any other x_{ℓ} , that there is a constant $c > 1$ such that

$$E[h(q)^2] \geq c^{-1} E[h_k(\tilde{q}_k)^2].$$

To show this property, note that holding fixed the vector of components $\tilde{q}_k^C$ of q that are not components of $\tilde{q}_k$ each $l \neq k$, $h_l(\tilde{q}_l)$ is a function of a strict subvector of $\tilde{q}_k$. Then,

$$\begin{aligned}
E[h(q)^2] &\geq c^{-1} \int \{h_k(\tilde{q}_k) + \sum_{l \neq k} h_l(\tilde{q}_l)\}^2 dF(\tilde{q}_k) dF(\tilde{q}_k^C) \\
&= c^{-1} \int [\int \{h_k(\tilde{q}_k) + \sum_{l \neq k} h_l(\tilde{q}_l)\}^2 dF(\tilde{q}_k)] dF(\tilde{q}_k^C) \\
&= c^{-1} \int [\int \{h_k(\tilde{q}_k)^2 + \{\sum_{l \neq k} h_l(\tilde{q}_l)\}^2\} dF(\tilde{q}_k)] dF(\tilde{q}_k^C) \\
&\geq c^{-1} \int [\int h_k(\tilde{q}_k)^2 dF(\tilde{q}_k)] dF(\tilde{q}_k^C) = c^{-1} E[h_k(\tilde{q}_k)^2].
\end{aligned}$$

To show the second conclusion, now add back the 2. subscript, and let $\mathcal{H} = \{q'_1 \beta + h_2(q_2) : h_2 \in \mathcal{H}_2\}$. Consider a sequence $h_j \in \mathcal{H}$, $h_j \rightarrow h_0$ in mean square. To show $h_0 \in \mathcal{H}$, note β_j is a mean-square continuous function of h_j , so that by h_j a Cauchy sequence, β_j is a Cauchy sequence, and hence converges to some β_0 , and hence $q'_1 \beta_j$ converges to $q'_1 \beta_0$ in mean-square, and hence $h_{2j} = h_j - q'_1 \beta_j$ converges to $h_0 - q'_1 \beta_0$, which is an element of $\mathcal{H}_2$ by $\mathcal{H}_2$ closed, so $h_0 \in \mathcal{H}$. ■

Lemma 8.1: If the support Q of q_i is a box and there is C such that $\mathcal{F} = \{f(q): \text{each additive component } \tilde{f}(q) \text{ of } f(q) \text{ is continuously differentiable of order } \ell \text{ and } \max_Q |\partial^\lambda \tilde{f}(q)| \leq C\}$, then for power series, there is $C > 0$ such that for all $f \in \mathcal{F}$, $\inf_{\pi \in \mathbb{R}^K} \|f(q) - p^K(q)' \pi\|_{d, \infty} < CK^{-\alpha}$. for $d = 0$ and $\alpha = \ell/r$, and for $r = 1$, $\ell < d$, and $\alpha = \ell - d$.

Proof: First, note that it suffices to show the result for $r = r$, since the approximation error of the function is bounded by the sum of errors over all additive components. For the first conclusion, note that by $|\lambda(K)|$ monotonic increasing, the set of all linear combinations of $p^K(q)$ will include the set of all polynomials of degree $CK^{1/r}$ for some C small

enough, so Theorem 8 of Lorentz (1986) applies. For the second conclusion, let $Q = [q^1, q^2]$, and note that $\partial^d p^{K+d}(q)/\partial q^d$ is a spanning vector for power series up to order K . By the first conclusion, there exists C such that for all $f \in \mathcal{F}$, there is π_K with $f_K(q) = p^{K+d}(q)' \pi_K$ satisfying $\sup_Q |\partial f(q)/\partial q - \partial f_K(q)/\partial q| \leq C \cdot K^{-\ell+d}$. The second conclusion then follows by integration and boundedness of Q , e.g. for $d = 1$ and the constant coefficient chosen so that $f(q_1) = f_K(q_1)$, $|f(q) - f_K(q)| \leq \int_{q_1}^q |\partial^d f(\tilde{q})/\partial q^d - \partial^d f_K(\tilde{q})/\partial q^d| d\tilde{q} \leq CK^{-\ell+1}$. ■

Lemma 8.2: If the support Q of q_i is star-shaped and there is C such that $\mathcal{F} = \{f(q): \text{each additive component } \tilde{f}(q) \text{ of } f(q) \text{ is continuously differentiable of all orders and for all multi-indices } \lambda, \max_Q |\partial^\lambda f(q)| \leq C^{|\lambda|}\}$, then for power series, for all $\alpha, d > 0$ there is $C > 0$ such that for all $f \in \mathcal{F}$, $\inf_{\pi \in \mathbb{R}^K} \|f(q) - p^K(q)' \pi\|_{d,\infty} \leq CK^{-\alpha}$.

Proof: As above, assume without loss of generality (w.l.g.) that $r = \iota$. By Q star-shaped, there exists $\bar{q} \in Q$ such that for all $q \in Q$, $\beta q + (1-\beta)\bar{q} \in Q$ for all $0 \leq \beta \leq 1$. For a function $f(q)$, let $P(f,m,q)$ denote the Taylor series up to order m for an expansion around $\bar{q}$. Note $\partial P(f,m,q)/\partial q_j = P(\partial f/\partial q_j, m-1, q)$, so that by induction $\partial^\lambda P(f,m,q) = P(\partial^\lambda f, m-|\lambda|, q)$. Also, $\partial^\lambda f(q)$ also satisfies the hypotheses, so that by the intermediate value form of the remainder,

$$\max_{q \in Q} |\partial^\lambda f(q) - P(\partial^\lambda f, m-|\lambda|, q)| \leq C^m / [(m-d)!].$$

Next, let $m(K)$ be the largest integer such that $P(f,m,q)$ is a linear combination of $p^K(q)$, and let $f_K(q) = P(f, m(K), q)$. By the "natural ordering" hypothesis, there are constants C_1 and C_2 such that $C_1 m(K)^r \leq K \leq C_2 m(K)^r$, so that for any $\alpha > 0$, $C^{m(K)} / [(m(K)-d)!] \leq CK^{-\alpha}$, and

$$\sup_{|\lambda| \leq d, Q} |\partial^\lambda f(q) - \partial^\lambda f_K(q)| = \sup_{|\lambda| \leq d, Q} |\partial^\lambda f(q) - P(\partial^\lambda f, m(K) - |\lambda|, q)| \leq CK^{-\alpha}. \quad \blacksquare$$

Lemma 8.3: If the support Q of q_i is a box, $r = 1$ or $d = 0$, there is C such that $\mathcal{F} = \{f(q): \text{each additive component } \tilde{f}(q) \text{ of } f(q) \text{ is continuously differentiable of order } \ell \text{ and } \max_Q |\partial^\lambda \tilde{f}(q)| \leq C\}$, and $m \geq \ell - 1$, then for splines there is $C > 0$ such that for all $f \in \mathcal{F}$, $\inf_{\pi \in \mathbb{R}^K} \|f(q) - p^K(q)' \pi\|_{d, \infty} < CK^{-\alpha}$, $r = 1$, $\ell < d$, and $\alpha = (\ell/r) - d$.

Proof: The result for $d = 0$ follows by Theorem 12.8 of Schumaker (1981).

For the other case, w.l.g. assume $r = 1$ and let $Q = [-1, 1]$. Note that $\partial^d p^K(q)/\partial q^d$ is a spanning vector for splines of degree $m - d$, with knot spacing bounded by CK^{-1} for K large enough and some C . Therefore, by Powell (1981), there exists π_K such that for $f_K(q) = p^K(q)' \pi_K$, $\sup_Q |\partial^d f(q)/\partial q^d - \partial^d f_K(q)/\partial q^d| \leq C \cdot K^{-\ell+d}$. The conclusion then follows by integration. $\blacksquare$

Lemma 8.4: If Assumption 6.1 is satisfied, then for power series Assumption 3.4 is satisfied and for any $d > 0$ there is a constant C such that for all $|\lambda| \leq d$, $\sup_{q \in Q, k \leq K} |\partial^\lambda P_{kK}(q)| \leq CK^{.5 + \bar{\nu} + 2|\bar{\lambda}|}$.

Proof: First, assume that q_1 is nonexistent, and let $q = q_2$. Following the definitions in Abramowitz and Stegun (1972, Ch. 22), let $C_k^{(\alpha)}(x)$ denote the ultraspherical polynomial of order k for exponent α , $h_k^{(\alpha)} \equiv \pi 2^{1-2\alpha} \Gamma(k+2\alpha) / \{k! (k+\alpha) [\Gamma(\alpha)]^2\}$, and $\rho_k^{(\alpha)}(x) = [h_k^{(\alpha)}]^{-1/2} C_k^{(\alpha)}(x)$. Also, let $x_j(q_j) = (2q_j - q_j^1 - q_j^2) / (q_j^2 - q_j^1)$ and define

$$P_k(q) = \prod_{j=1}^r \rho_{\lambda_j(k)}^{(\nu_j+.5)}(x_j(q_j)).$$

$P^K(q)$ is a nonsingular combination of $p^K(q)$ by the "natural ordering"

assumption (i.e. by $|\lambda(k)|$ monotonic increasing). Also, for $\tilde{P}(q)$ absolutely continuous on $Q = \prod_{j=1}^r [q_j^1, q_j^2]$ with pdf proportional to $\prod_{j=1}^r [(q_j^2 - q_j)(q_j - q_j^1)]^{\nu_j}$, and by the change of there is a constant C with

$$\lambda_{\min}(\int P^K(q) P^K(q)' d\tilde{P}(q)) \\ \geq \lambda_{\min}(\int \otimes_{j=1}^r [\tilde{\rho}_M^{(\nu_j + .5)}(x_j(q_j)) \tilde{\rho}_M^{(\nu_j + .5)}(x_j(q_j))'] d\tilde{P}(q)) = C,$$

where the inequality follows by $P^K(q)$ a subvector of $\otimes_{j=1}^r [\tilde{\rho}_M^{(\nu_j + .5)}(x_j(q_j))]$ for $M = \max_{k \leq K} |\lambda(k)|$ and $\tilde{p}_M^{(\alpha)}(x) = (p_0^{(\alpha)}(x), \dots, p_M^{(\alpha)}(x))$.

Next, by differentiating 22.5.37 of Abramowitz and Stegun (for m there equal to ν here) and solving, it follows that for $\ell \leq k$, $d^\ell C_k^{(\nu + .5)}(x)/dx^\ell = C \cdot C_{k-\ell}^{(\nu + \ell + .5)}(x)$ so that by 22.14.2 of Abramowitz and Stegun, for $\lambda(k-s)$ as in equation (2.3),

$$|\partial^{\lambda} P_{kK}(q)| \leq C \prod_{j=1}^r [1 + \lambda_j(k-s)]^{.5 + \nu_j + 2\lambda_j} \leq C |\lambda(k-s)|^{n(.5 + \bar{\nu} + \bar{\lambda})} \leq C \cdot K^{.5 + \bar{\nu} + 2\bar{\lambda}},$$

where the last equality follows by $|\lambda(k-s)| \leq CK^{1/n}$.

Now, for the case with q_1 , let $P_{kK}(q) = q_{1k}$, $k = 1, \dots, s$, $P_{kK}(q) = P_{k-s}(q)$ for $P_k(q)$ described above. The bounds of the previous equation continue to hold by q_1 bounded. Let $P(q_2) = (P_{s+1}(q_2), \dots, P_K(q_2))'$ and $\mathcal{H}_{2K}$ the linear space spanned $P(q_2)$. Note $\mathcal{H}_{2K} \subseteq \mathcal{H}_2$, so that $M_K E[\{q_1 - P(q_1 | \mathcal{H}_{2K})\} \{q_1 - P(q_1 | \mathcal{H}_{2K})\}']$ is bigger than M , and thus has smallest eigenvalue bounded away from zero by Assumption 6.1. Furthermore, $E[P^K(q) P^K(q)'] = BDB'$ for,

$$B = \begin{bmatrix} I & E[q_1 P(q_2)'] (E[P(q_2) P(q_2)'])^{-1} \\ 0 & I \end{bmatrix}, \quad D = \begin{bmatrix} M_K & 0 \\ 0 & E[P(q_2) P(q_2)'] \end{bmatrix}.$$

Thus, by the extremal characterization of the smallest eigenvalue,

$$E[P^K(q)P^K(q)'] \geq \lambda_{\min}(BB')\lambda_{\min}(D) \geq \min\{\lambda_{\min}(M_K), \lambda_{\min}(E[P(q_2)P(q_2)'])\} \geq C. \blacksquare$$

Lemma 8.5: If Assumption 7.1 is satisfied then for splines Assumption 3.4 is satisfied, the number of nonzero elements of $P^K(q)P^K(q)'$ is bounded by CK , and for any $d > 0$ there is a constant C such that for all λ with $|\lambda| \leq d$ and $\lambda_j < m$, ($j = 1, \dots, r$), it follows that $\sup_{q \in Q, k \leq K} |\partial^\lambda P_{kK}(q)| \leq CK^{5+|\bar{\lambda}|}$.

Proof: First, consider the case where $q = q_2$ and let $Q = \prod_{\ell=1}^r [-1, 1]$. Let $B_{jL}^m(x)$, be the B-spline of order m , for the knot sequence $-1 + 2j/[L+1]$, $j = \dots, -1, 0, -1, \dots$ with left end-knot j , and let

$$P_{\ell k}(q_\ell) \equiv (L_\ell/2)^{1/2} B_{k-m-1, L_\ell}^m(q_\ell), \quad (k = 1, \dots, L_\ell+m+1, \ell = 1, \dots, r),$$

$$P_{kK}(q) = \prod_{\ell=1}^r 1(\lambda_\ell(k) > 0) P_{\ell, \lambda_\ell(k)}(q_\ell).$$

Then existence of a nonsingular matrix A such that $P^K(q) = AP^K(q)$ for $q \in Q$ follows by inclusion in $P^K(q)$ of all multiplicative interactions of splines for components of q corresponding to components of $h(q)$ and the usual basis result for B-splines (e.g. Theorem 19.2 of Powell, 1981).

Next, by a well known property of B-splines, $P_{\ell k}(q_\ell)P_{\ell k'}(q_\ell) = 0$ for all $q \in \mathbb{R}$ if $|k-k'| > m$, implying that the number of nonzero elements of $P^K(q)P^K(q)'$ is bounded above by $2(m+1)^r K = CK$. Also, for $\tilde{P}(q)$ the distribution of r independent uniform random variables on Q , noting that $[2(m+1)/L_\ell][L_\ell/2]^{-1/2} P_{\ell k}(q_\ell)$ are the so-called normalized B-splines with evenly spaced knots, it follows by the argument of Burman and Chen (1989, p. 1587) that for $P_{\ell, L}(q) = (P_{\ell 1}(q_\ell), \dots, P_{\ell, L+m+1}(q_\ell))$, there is C with $\lambda_{\min}(\int_{-1}^1 P_{\ell, L}(q)P_{\ell, L}(q)'dq) \geq C$ for all positive integers L . Therefore, the boundedness away from zero of the smallest eigenvalue follows by $P^K(q)$ a

subvector of $\otimes_{\ell=1}^r P_{\ell, L_\ell}(q_\ell)$, analogously to the proof of Lemma 8.4. Also, since changing even knot spacing is equivalent to rescaling the argument of B-splines, $\sup_{\mathbb{R}} |\partial^d B_{jL}^m(x)/\partial x^d| \leq CL^d$, $d \leq m$, implying the bounds on derivatives given in the conclusion. The proof when q_1 is present follows as in the proof of Lemma 8.4 ■

Proof of Theorem 4.1: Note that Assumptions 3.2 and 3.3 imply that Assumption 9.1 is satisfied for $J = 1$ and $y_1 = y$. The first conclusion then follows by Lemma 9.9, and the second by Lemma 9.10. ■

Proof of Theorem 4.2: By reasoning as in the previous proof, the theorem immediately from Lemma 9.11. ■

Proof of Theorem 4.3: Follows immediately from Theorem 4.2. ■

Proof of Theorem 5.1: By Assumption 3.4, $P^K(q)$ is a nonsingular linear combination of $p^K(q)$, so that replacing $p^K(q)$ by $P^K(q)$ does not change $A(\hat{h})$, Σ , $\hat{\Omega}$, or Ω . Thus, it suffices to show the conclusion with this replacement. Note that $\|DB\| = [\text{tr}(DB^2D')]^{1/2} \leq \|D\|_{\lambda_{\max}}(B^2)^{1/2} = \|D\|_{\lambda_{\max}}(B)$ for any matrix D and positive semi-definite B . Thus, for $F = \Omega^{-1/2}$, by $\sigma^2(q) \geq C$, for a positive definite square-root $\Sigma^{-1/2}$,

$$(8.1) \quad \|FA'\Sigma^{-1/2}\| = \{\text{tr}\{FA'\Sigma^{-1}AF'\}\}^{1/2} \leq C\{\text{tr}\{FA'\Sigma^{-1}V\Sigma^{-1}AF'\}\}^{1/2} \leq C\sqrt{n},$$

$$\|FA'\Sigma^{-1}\| \leq \|FA'\Sigma^{-1/2}\|_{\lambda_{\max}}(\Sigma^{-1/2}) \leq C\sqrt{n}.$$

Note that $A'\Sigma^{-1}A$ is invariant to replacing $P^K(q)$ by the $C_K P^K(q)$ from Assumption 5.2, so that in analyzing the properties of $A'\Sigma^{-1}A$ the j^{th} element of $P^K(q)$, and hence the j^{th} row of A , is invariant to K . It then follows that $A'\Sigma^{-1}A$ is a monotonic increasing sequence in the positive semi-definite semi-order (since this matrix is formally identical to the

inverse asymptotic variance of a minimum chi-square estimator, which rises as additional equality restrictions are added). Therefore, the smallest eigenvalue of $A'\Sigma^{-1}A$ is also monotonic increasing in K , so that by A having full rank for some K , the smallest eigenvalue of $A'\Sigma^{-1}A$ is bounded away from zero. This implies that

$$(8.2) \quad \|F\| \leq C\lambda_{\max}(\Omega^{-1})^{1/2} \leq C\lambda_{\min}(\Omega)^{-1/2} \leq C\sqrt{n}.$$

Also, $\hat{\Sigma}^{-1}-\Sigma^{-1} = \hat{\Sigma}^{-1}(\Sigma-\hat{\Sigma})\Sigma^{-1} = \Sigma^{-1}(\Sigma-\hat{\Sigma})\Sigma^{-1} + (\hat{\Sigma}^{-1}-\Sigma^{-1})(\Sigma-\hat{\Sigma})\Sigma^{-1} = \Sigma^{-1}(\Sigma-\hat{\Sigma})\Sigma^{-1} + \Sigma^{-1}(\Sigma-\hat{\Sigma})\hat{\Sigma}^{-1}(\Sigma-\hat{\Sigma})\Sigma^{-1}$, by Lemma 9.6 and Assumption 3.7, $\|\hat{\Sigma}-\Sigma\| = o_p(1)$, $\lambda_{\max}(\hat{\Sigma}^{-1}) = O_p(1)$, and there are positive semidefinite square roots $\Sigma^{-1/2}$, $\hat{\Sigma}^{-1/2}$ such that $\lambda_{\max}(\Sigma^{-1/2}) = O(1)$ and $\lambda_{\max}(\hat{\Sigma}^{-1/2}) = O_p(1)$. It follows that $\|A\Sigma^{-1}\| \leq C\|A\Sigma^{-1/2}\|$, so that w.p.a.1,

$$\|FA'\hat{\Sigma}^{-1/2}\|^2 \leq Cn[1 + \lambda_{\max}(\Sigma^{-1/2})\|\Sigma-\hat{\Sigma}\|(1 + \|\Sigma-\hat{\Sigma}\|O_p(1))] = O_p(n),$$

$$\|FA'\hat{\Sigma}^{-1}\|^2 \leq \|FA'\hat{\Sigma}^{-1/2}\|^2 \lambda_{\max}(\hat{\Sigma}^{-1/2})^2 = O_p(n),$$

Let $\tilde{\pi}$ be such that $|h(q) - P^K(q)'\tilde{\pi}|_{d,v} \leq CK^{-\alpha}$. For $\tilde{h} = (h(q_1), \dots, h(q_n))'$,

$$F[A(\hat{h}) - A(h_0)] = FA'\hat{\Sigma}^{-1}P'u/n + R, \quad R = F[A'\hat{\Sigma}^{-1}P'(\tilde{h}-P\tilde{\pi})/n + A'\tilde{\pi} - A(h_0)],$$

where, for convenience, the K superscript on P^K has been dropped. By vi) and (8.2), w.p.a.1,

$$(8.3) \quad \begin{aligned} \|R\| &\leq \|FA'\hat{\Sigma}^{-1}P/\sqrt{n}\|\|\tilde{h}-P\tilde{\pi}\|/\sqrt{n} + \|F\|\|A'\tilde{\pi} - A(h_0)\| \\ &\leq C[\text{tr}(FA'\hat{\Sigma}^{-1}\hat{\Sigma}\hat{\Sigma}^{-1}AF')]^{1/2} + \sqrt{n}K^{-\alpha} \leq C\sqrt{n}K^{-\alpha} = o_p(1). \end{aligned}$$

Also, $FA'\hat{\Sigma}^{-1}P'u/n = FA'\Sigma^{-1}P'u/n + R$, $R = FA'(\hat{\Sigma}^{-1}-\Sigma^{-1})P'u/n$. By the proof of Lemma 9.8, $\|\Sigma^{-1/2}P'u/\sqrt{n}\| = O_p(K^{1/2})$ and by Lemma 9.6, $\|(\Sigma-\hat{\Sigma})\Sigma^{-1/2}\| \leq$

$\|\Sigma - \hat{\Sigma}\|_{\lambda_{\max}}(\Sigma^{-1/2}) = o_p(\tilde{K}^{1/2}\zeta_0(K)^2/\sqrt{n})$. Therefore,

$$(8.4) \quad \begin{aligned} \|R\| &= \|FA'\hat{\Sigma}^{-1}(\Sigma - \hat{\Sigma})\Sigma^{-1}P'u/n\| \leq \|FA'\hat{\Sigma}^{-1}/\sqrt{n}\| \|(\Sigma - \hat{\Sigma})\Sigma^{-1/2}\| \|\Sigma^{-1/2}P'u/\sqrt{n}\| \\ &\leq CO_p(K^{1/2}\tilde{K}^{1/2}\zeta_0(K)^2/\sqrt{n}) = o_p(1). \end{aligned}$$

Next, let v be a constant vector with $\|v\| = 1$ and the same dimension as $A(h)$, and $Z_{in} \equiv v'FA'\Sigma^{-1}P^K(q_i)u_i/\sqrt{n}$, so that $\sum_{i=1}^n Z_{in}/\sqrt{n} = v'FA'\Sigma^{-1}P'u/n$. Note that Z_{in} is a martingale difference sequence, and $E[Z_{in}^2] = 1$. Also, $|Z_{in}| \leq \|v\| \|FA'\Sigma^{-1}\| \|P^K(q_i)\| |u_i|/\sqrt{n} \leq CK^{1/2}\zeta_0(K)|u_i|$. By Assumption 5.1, for $\Delta = s/2$, $\mu[1-(2/\Delta)] > 1$ and $|u_i^2|_{\Delta} < \infty$. Thus, $(|Z_{in}^2|_{\Delta})^2 \leq CK^2\zeta_0(K)^4(|u_i^2|_{\Delta})^2 \leq CK^2\zeta_0(K)^4$. Then by Davydov's (1968) inequality and $K^2 \leq K\tilde{K}$,

$$E[(\sum_{i=1}^n Z_{in}^2/n - 1)^2] \leq (C/n)[\sum_{t=0}^{n-1} t^{-\mu(1-2/\Delta)}] (|Z_{in}^2|_{\Delta})^2 \leq CK^2\zeta_0(K)^4/n = o(1),$$

so that $\sum_{i=1}^n Z_{in}^2/n \xrightarrow{p} 1$. Also, for any $\epsilon > 0$,

$$E[1(Z_{in}^2 > \epsilon \cdot n)Z_{in}^2] \leq E[Z_{in}^4]/(n\epsilon) \leq CK^2\zeta_0(K)^4 E[|u_i|^4]/(n\epsilon) = o(1).$$

It then follows by Theorem 5.2.3 of White (1984) that $\sum_{i=1}^n Z_{in}/\sqrt{n} \xrightarrow{d} N(0, 1)$. Since this result holds for all v with $\|v\| = 1$, $FA'\Sigma^{-1}P'u/n \xrightarrow{d} N(0, I)$ follows by the Cramer-Wold device. The first conclusion then follows by eqs. (8.3), (8.4), and the triangle inequality.

To prove the second conclusion, note that Assumption 5.1 implies that the hypotheses of Theorem 4.1 are satisfied. Thus, for $\hat{u}_i = y_i - \hat{h}(q_i)$, $\sum_{i=1}^n |\hat{u}_i - u_i|^2/n = \sum_{i=1}^n |\hat{h}(q_i) - h(q_i)|^2/n = O_p(K/n + (nK^{-2\alpha})/n) = O_p(K/n) = o_p(1)$. Therefore, by $|\hat{u}_i^2 - u_i^2| \leq 2|u_i||\hat{u}_i - u_i| + |\hat{u}_i - u_i|^2$,

$$(8.5) \quad \sum_{i=1}^n |\hat{u}_i^2 - u_i^2|/n \leq O_p(1) \{ \sum_{i=1}^n |\hat{u}_i - u_i|^2/n \}^{1/2} = O_p(K^{1/2}/n^{1/2}).$$

Let $\hat{V}$ and $\hat{\Sigma}$ be as defined above, except with $P^K(q)$ replacing $p^K(q)$, and let $\tilde{V} \equiv \sum_i P^K(q_i) P^K(q_i)' u_i^2/n$. Then

$$(8.6) \quad \|\hat{V} - \tilde{V}\| \leq \tilde{K}^{1/2} \zeta_0(K)^2 \sum_i |\hat{u}_i^2 - u_i^2|/n = o_p(\tilde{K}^{1/2} \zeta_0(K)^2 K^{1/2}/n^{1/2}) = o_p(1).$$

Next, apply Lemma 9.6, with $y_j = P_{KK}(q_1) P_{LK}(q_1) u_1^2$, noting that by $s > 4\mu/(\mu-1)$, Assumption 9.1 is satisfied for $J = \tilde{K}$, $v_y(J) = \zeta_0(\tilde{K})^2$, $B_{y1} = u_1^2$, and s there equal to $2s$, so that $\|\tilde{V} - V\| = o_p(\tilde{K}^{1/2} \zeta_0(\tilde{K})^2/\sqrt{n}) = o_p(1)$. It now follows by the triangle inequality that $\|\hat{V} - V\| = o_p(1)$. Therefore,

$$\begin{aligned} \|F[\hat{\Omega} - \Omega]F'\| &\leq \|FA'\hat{\Sigma}^{-1}\|^2 \|\hat{V} - V\| + \|F[A'\hat{\Sigma}^{-1}(\Sigma - \hat{\Sigma})\Sigma^{-1}V\Sigma^{-1}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-1}A]F'\| \\ &\quad + 2\|F[A'\hat{\Sigma}^{-1}(\Sigma - \hat{\Sigma})\Sigma^{-1}V\Sigma^{-1}A]F'\| \\ &\leq o_p(1) + o_p(1)\lambda_{\max}(\Sigma^{-1}) + o_p(1)\lambda_{\max}(I) = o_p(1), \end{aligned}$$

giving the second conclusion. If the final hypothesis is satisfied, then $\psi_n^{1/2} \Omega^{1/2} \rightarrow \Omega_0^{1/2}$ by continuity of the square root and $(\psi_n \Omega)^{1/2} = \psi_n^{1/2} \Omega^{1/2}$. Therefore, multiplying through by $\psi_n^{1/2} \Omega^{1/2}$, it follows that $\psi_n \hat{\Omega} \xrightarrow{p} \Omega_0$, and, by Ω_0 nonsingular, that $\psi_n^{-1/2} \hat{\Omega}^{-1/2} \xrightarrow{p} \Omega_0^{-1/2}$. Therefore, $\hat{\Omega}^{-1/2} \Omega^{1/2} \xrightarrow{p} I$, so the final conclusion follows from the first conclusion.

Proof of Theorem 5.2: By iii) and each component of $p^K(q)$ an element of $\mathcal{H}$, $A = E[p^K(q)\delta(q)']$. Let $\delta_K(q) \equiv p^K(q)'\Sigma^{-1}A = p^K(q)'\Sigma^{-1}E[p^K(q)\delta(q)']$. Since $\delta_K(q)$ is the minimum mean-square error linear combination of $p^K(q)$, it follows by ii) and $\sigma^2(q)$ bounded that $E[\|\delta(q) - \delta_K(q)\|^2] \rightarrow 0$ and $E[\sigma^2(q)\|\delta(q) - \delta_K(q)\|^2] \leq CE[\|\delta(q) - \delta_K(q)\|^2] \rightarrow 0$. Therefore, $A'\Sigma^{-1}A = E[\delta_K(q)\delta_K(q)'] \rightarrow E[\delta(q)\delta(q)']$, so that Assumption 5.2 is satisfied by iii). Also,

$$\|n\Omega - \Omega_0\| = \|E[\sigma^2(q)A'\Sigma^{-1}p^K(q)p^K(q)'\Sigma^{-1}A] - \Omega_0\|$$

$$= \|E[\sigma^2(q)\delta_K(q)\delta_K(q)'] - E[\sigma^2(q)\delta(q)\delta(q)']\| \leq E[\sigma^2(q)\|\delta(q) - \delta_K(q)\|^2] \rightarrow 0,$$

so that the final hypothesis of Theorem 5.1 is satisfied. The conclusion follows by the final conclusion of Theorem 5.1. ■

Proof of Theorem 5.3: If $A(h)$ is not mean square continuous then there exists a sequence $h_j(q) \in \mathcal{H}$, ($j = 1, 2, \dots$) such that $E[h_j(q)^2] \rightarrow 0$ and $|A(h_j)|$ is bounded away from zero. Consider any parametric submodel such that $P(y|\mathcal{H}) = h(q) + \gamma h_j(q)$, with true value of γ equal to zero. By Chamberlain (1987) the supremum over all such submodels of the Cramer-Rao variance bound is the asymptotic variance of the least squares estimator, which is $(E[h_j(q)^2])^{-2}E[\sigma^2(q)h_j(q)^2]$. Furthermore, by the delta-method and $\sigma^2(q)$ bounded away from zero, the corresponding supremum for $A(h)$ is

$$[\partial A(h + \gamma h_j) / \partial \gamma]^2 (E[h_j(q)^2])^{-2} E[\sigma^2(q)h_j(q)^2] \geq CA(h_j)^2 (E[h_j(q)^2])^{-1} \rightarrow \infty.$$

Therefore, the supremum over all parametric submodels of Cramer-Rao bounds for $A(h)$ is not finite. ■

Proof of Theorem 5.4: Given in Section 5. ■

Proof of Theorem 5.5: Given in Section 5. ■

Proof of Theorem 6.1: Proceed by verifying the hypotheses of Theorems 4.1 and 4.2. Assumption 3.4 with $\zeta_0(K) = K^{5+\bar{\nu}}$ follows by Assumption 6.1 and Lemma 8.4. Assumptions 3.5 and 3.6 follow by the assumption $|\lambda(k-s)|$ is increasing, which implies that the products of univariate orthogonal polynomial terms form a nested sequence. Assumption 3.7 follows by Assumption 6.2 and $\tilde{K} = K^2$. Assumption 3.8 with $d = 0$ follows by Lemma 8.1. The first two conclusions then follow from Theorem 4.1. The third line follows from

Theorem 4.2 similarly. The final conclusion follows from Theorem 4.2, the bound on $\zeta_\lambda(K)$ from Lemma 8.4, and Lemma 8.2 (which implies Assumption 8 for any $\alpha > 0$). ■

Proof of Theorem 6.2: By Theorem 5.1 it suffices to show that Assumption 5.3 is satisfied. Assumption 3.4 with $\zeta_0(K) = K^{5+\bar{\nu}}$ follows by Assumption 6.1 and Lemma 8.4. Assumptions 3.8 and 3.9 and $\sqrt{n}K^{-\alpha} \rightarrow 0$ follow by ii) and Lemmas 8.1 and 8.2. Finally, note that $\tilde{K} = K^2$, so that $K\tilde{K}\zeta_0(K)^{4/n} = K^3K^{2+4\bar{\nu}/n} = K^{5+4\bar{\nu}/n} \rightarrow 0$ by i). ■

Proof of Theorem 6.3: Proceed to verify hypotheses of Theorem 6.3.

Assumptions 3.1, 3.2, and 5.1 are satisfied by the independence of the observations. Note that $A(h) = h_j(\tilde{q}_j) - h_j(\bar{q}_j) = h(\bar{q}_1, \dots, \tilde{q}_j, \dots, \bar{q}_r) - h(\bar{q})$, is continuous with respect to $|h|_{0,\omega}$, while $A(h) = \partial^d h_j(\tilde{q}_j) / \partial q_j^d = \partial^d h(\tilde{q}) / \partial q_j^d$ is continuous with respect to $|h|_{d,\omega}$, so that ii) of Theorem 6.3 is satisfied. The conclusion then follows by taking $\psi_n = \Omega^{-1/2}$. ■

Proof of Theorem 6.4: Proceed from Theorem 5.2. It follows by $A(\tilde{h}) = E[\delta(q)\tilde{h}(q)]$, as in the proof of Theorem 6.3, that Assumption 5.3 is satisfied with $v = 2$. Theorem 5.2 ii) follows by the well known spanning result for power series for bounded q (e.g. Gallant, 1980), giving the result. ■

Proof of Theorem 6.5: Follows from Theorem 5.3 by the same argument used in the proof of Theorem 6.5. ■

Proof of Theorem 7.1: Proceed by verifying the hypotheses of Theorems 4.1 and 4.2. Assumption 3.4 with $\zeta_0(K) = K^5$ follows by Assumption 7.1 and Lemma 8.5. Assumptions 3.5 and 3.6 follow trivially by K constant. Assumption 3.7 follows by Assumption 7.2 and Lemma 8.5, which implies $\tilde{K} \leq CK^2$. Assumption 3.8 with follows by Lemma 8.3. The conclusions then follow from

Theorems 4.1 and 4.2, with the bound on $\zeta_\lambda(K)$ from Lemma 8.5. ■

Proof of Theorem 7.2: By Theorem 5.1 it suffices to show that Assumption 5.3 is satisfied. Assumption 3.4 with $\zeta_0(K) = K^5$ follows by Assumption 7.1 and Lemma 8.5. Assumptions 3.8 and 3.9 and $\sqrt{n}K^{-\alpha} \rightarrow 0$ follow by ii) and Lemma 8.3. Finally, note that $\tilde{K} \leq CK$ by Lemma 8.5 so that $K\tilde{K}\zeta_0(K)^4/n = K^2K^2/n = K^4/n \rightarrow 0$ by i). ■

Proof of Theorem 7.3: Follows analogously to the proof of Theorem 6.3. ■

Proof of Theorem 7.4: Follows analogously to the proof of Theorem 6.5, except for the explicit formula for $\delta(q)$ given, which follows by

$$\int w(q_j) [\partial^d h_j(q_j) / \partial q_j^d] dq_j = \int w(q) [\partial^d h(q) / \partial q_j^d] dq \text{ for } w(q) = w(q_j) f(q_{-j}), \text{ where} \\ [\partial^d w(q) / \partial q_j^d] / f(q) = [\partial^d w(q_j) / \partial q_j^d] f(q_{-j}) / f(q) = [\partial^d w(q_j) / \partial q_j^d] / f(q_j | q_{-j}).$$

9. Useful Lemmas

This Section gives general results on convergence rates for certain remainder terms. It is useful to allow throughout for a vector of series estimates with dimension that can increase with sample size. To do so, it is necessary to introduce more notation and assumptions.

Let $\{y_{jJ}\}_{j=1, J=1}^J, \infty$ be a collections of functions of a single data observation z . For notational convenience, the J subscript on y_{jJ} will be suppressed in what follows. The results will pertain to certain of sample averages of these functions, or of series estimators of the projections $h_j(q)$ of y_j on $\mathcal{H}$. Denote the observations on y_j by y_{ij} , and let $u_{ij} \equiv y_{ij} - h_j(q_i)$ and objects without a subscript denote corresponding vectors of n

or J observations, e.g. $y_j = (y_{1j}, \dots, y_{nj})'$ and $u_i = (u_{i1}, \dots, u_{iJ})'$. For $\hat{K} = \hat{K}(z_1, \dots, z_n, n)$, let $p \equiv [p^{\hat{K}}(q_1), \dots, p^{\hat{K}}(q_n)]'$. The estimators are $\hat{h}_j(q) = p^{\hat{K}}(q)'(p'p)^{-1}p'y_j$.

Assumption 9.1: For $s > 1$ and $\nu_y(J)$, $\max_{j \leq J} |u_{ij}| \leq \nu_y(J)B_{yi}$, $|B_{yi}|_s < \infty$, $E[B_{yi}^2 | q_i] \leq C$ and either a) z_t is uniform (ϕ) mixing with mixing coefficients $\phi(t) = O(t^{-\mu})$, $(t = 1, 2, \dots)$, $\mu > 2$, or b) there exists $\gamma(t)$ such that $\sum_{t=1}^{\infty} \gamma(t) < \infty$ and $\max_{j \leq J} |E[u_{ij}u_{i+j} | q_i, q_{i+j}]| \leq \gamma(t)\nu_y(J)^2$, $(t=1, 2, \dots)$.

Henceforth, let $\sum_{i=1}^n = \sum_i$, $\sum_{\ell=1}^r = \sum_{\ell}$, and $\sum_{j=1}^J = \sum_j$.

The first few Lemmas consist of useful convergence results for random matrices with dimension that can depend on sample size. Let $\hat{\Sigma}$ and Σ denote symmetric matrices such matrices, and $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ the smallest and largest eigenvalues respectively.

Lemma 9.1: If $\lambda_{\min}(\Sigma) \geq C$ with probability approaching one (w.p.a.1) and $\|\hat{\Sigma} - \Sigma\| = o_p(1)$ then $\lambda_{\min}(\hat{\Sigma}) \geq C$ w.p.a.1.

Proof: For a conformable vector μ , it follows by $\|\cdot\|$ a matrix norm that $\lambda_{\min}(\hat{\Sigma}) = \min_{\|\mu\|=1} \{\mu'\Sigma\mu + \mu'(\hat{\Sigma} - \Sigma)\mu\} \geq \lambda_{\min}(\Sigma) - \lambda_{\max}(\hat{\Sigma} - \Sigma) \geq \lambda_{\min}(\Sigma) - \|\hat{\Sigma} - \Sigma\| \geq C - o_p(1)$. Therefore, $\lambda_{\min}(\hat{\Sigma}) \geq C/2$ w.p.a.1. ■

Lemma 9.2: If $\lambda_{\min}(\Sigma) \geq C$ w.p.a.1, $\|\hat{\Sigma} - \Sigma\| = o_p(1)$, and D_n is a conformable matrix such that $\|\Sigma^{-1/2}D_n\| = O_p(\epsilon_n)$ for some ϵ_n , then $\|\hat{\Sigma}^{-1/2}D_n\| = O_p(\epsilon_n)$.

Proof: It is easy to show that for any conformable matrices A and B , $\|AB\| \leq \|A\| \cdot \|B\|$, $\|A'BA\| \leq \|B\| \cdot \|A'A\|$, and that if B is positive semi-definite, $\text{tr}(A'BA) \leq \|A\|^2 \lambda_{\max}(B)$, $\|AB\| \leq \|A\| \lambda_{\max}(B)$ and $\|BA\| \leq \|A\| \lambda_{\max}(B)$. Let

$\Sigma^{-1/2}$ be the symmetric square root of Σ^{-1} which is equal to $U\Lambda U'$ where U is an orthogonal matrix and Λ a diagonal matrix consisting of the square roots of the eigenvalues of Σ^{-1} . Note that $\Sigma^{-1/2}$ is positive definite and $\lambda_{\max}(\Sigma^{-1/2}) = [\lambda_{\max}(\Sigma^{-1})]^{1/2}$. Also by Lemma 9.1, $\lambda_{\max}(\hat{\Sigma}^{-1}) = O_p(1)$. Then

$$\begin{aligned}
 (9.1) \quad & \|\hat{\Sigma}^{-1/2} D_n\|^2 = \text{tr}(D_n' [\hat{\Sigma}^{-1} - \Sigma^{-1}] D_n) \\
 & \leq \|\Sigma^{-1/2} D_n\|^2 (1 + \|\Sigma^{-1/2} [\Sigma - \hat{\Sigma}] \Sigma^{-1/2}\|) + \|(\Sigma - \hat{\Sigma}) \Sigma^{-1} D_n\|^2 \lambda_{\max}(\hat{\Sigma}^{-1}) \\
 & \leq O_p(\epsilon_n^2) [1 + o_p(1) O_p(1) + \|\Sigma - \hat{\Sigma}\|^2 \lambda_{\max}(\Sigma^{-1/2})^2 O_p(1)] = O_p(\epsilon_n^2). \quad \blacksquare
 \end{aligned}$$

Let $\text{tr}(A)$ denote the trace of a square matrix A and u a random matrix with n rows.

Lemma 9.3: Suppose $\lambda_{\min}(\Sigma) \geq C$ w.p.a.1., $\bar{P}$ is a $\bar{K} \times n$ random matrix such that $\|\bar{P}'\bar{P}/n - \Sigma\| = o_p(1)$ and $\|\Sigma^{-1/2}\bar{P}'u/n\| = O_p(\epsilon_n)$, and $p = \bar{P}A$ where A is a random matrix. Then $\text{tr}(u'p(p'p)^{-1}p'u/n) = O_p(\epsilon_n^2)$.

Proof: Let $\bar{W} = \bar{P}(\bar{P}'\bar{P})^{-1}\bar{P}'$ and $W = p(p'p)^{-1}p'$ be the orthogonal projection operators for the linear spaces spanned by the columns of $\bar{P}$ and p respectively. Since the space spanned by p is a subset of the space spanned by $\bar{P}$, $\bar{W} - W$ is positive semi-definite. Let $\hat{\Sigma} = \bar{P}'\bar{P}/n$. Then by Lemma 9.2, $\text{tr}(u'Wu/n) \leq \text{tr}(u'\bar{W}u/n) = \|\hat{\Sigma}^{-1/2}\bar{P}'u/n\|^2 = O_p(\epsilon_n^2)$. $\blacksquare$

Let Y and G denote random matrices with the same number of columns and n rows, and let $u = Y - G$. For a vector p let $\hat{\pi} = (p'p)^{-1}p'Y$ and $\hat{G} = p\hat{\pi}$.

Lemma 9.4: If $\text{tr}(u'p(p'p)^{-1}p'u/n) = O_p(\epsilon_n^2)$. Then for any conformable matrix π , $\|\hat{G} - G\|^2/n \leq O_p(\epsilon_n^2) + \|G - p\pi\|^2/n$.

Proof: For W and $\bar{W}$ as in the proof of Lemma 9.3, by $Wp = p$, and $I - W$ idempotent,

$$\begin{aligned}\|\hat{G}-G\|^2/n &= \text{tr}[Y'WY - Y'WG - G'WY + G'G]/n = \text{tr}[u'Wu + G'(I-W)G]/n \\ &\leq \text{tr}[u'Wu + (G-p\pi)'(I-W)(G-p\pi)]/n \leq O_p(\epsilon_n^2) + \|G-p\pi\|^2/n. \quad \blacksquare\end{aligned}$$

Lemma 9.5: Suppose $\lambda_{\min}(\Sigma) \geq C$ w.p.a.1., $\bar{P}$ is a $\bar{K} \times n$ random matrix such that $\|\bar{P}'\bar{P}/n - \Sigma\| = o_p(1)$ and $\text{tr}[u'\bar{P}(\bar{P}'\bar{P})^{-1}\bar{P}'u/n] = O_p(\epsilon_n^2)$, and $P = \bar{P}S$ where S is a random selection matrix. Then for any conformable matrix π ,

$$\begin{aligned}\|\hat{\pi}-\pi\|^2 &\leq O_p(\epsilon_n^2) + O_p(1)\|G-p\pi\|^2/n, \\ \text{tr}[(\hat{\pi}-\pi)'S'\Sigma S(\hat{\pi}-\pi)] &\leq O_p(\epsilon_n^2) + O_p(1)\|G-p\pi\|^2/n.\end{aligned}$$

Proof: By $p'p/n = S'(\bar{P}'\bar{P}/n)S$ for the selection matrix S , $\lambda_{\min}(p'p/n) \geq \lambda_{\min}(\bar{P}'\bar{P}/n)$. Thus, $\lambda_{\min}(p'p/n) \geq C$ w.p.a.1, so $\lambda_{\min}(p'p/n)^{-1} = O_p(1)$. Also, note that for $\bar{W}$ as above, $W = p(p'p)^{-1}p'$, and $\tilde{G} = p\pi$,

$$\begin{aligned}\|\hat{\pi}-\pi\|^2 &\leq \lambda_{\min}(p'p/n)^{-1} \text{tr}[(\hat{\pi}-\pi)'(p'p/n)(\hat{\pi}-\pi)] \\ &= O_p(1) \text{tr}[Y'WY - Y'W\tilde{G} - \tilde{G}'WY + \tilde{G}'\tilde{G}]/n \\ &\leq O_p(1) [\text{tr}(u'Wu/n) + \|\tilde{G}-G\|^2/n] = O_p(\epsilon_n^2) + O_p(1)\|\tilde{G}-G\|^2/n.\end{aligned}$$

To prove the second conclusion, note that by the triangle inequality and the same arguments as for the previous equation,

$$\begin{aligned}\text{tr}[(\hat{\pi}-\pi)'S'\Sigma S(\hat{\pi}-\pi)] &= \text{tr}[(\hat{\pi}-\pi)'[S'\Sigma S - p'p/n](\hat{\pi}-\pi)] + (\hat{\pi}-\pi)'(p'p/n)(\hat{\pi}-\pi) \\ &\leq \|\hat{\pi}-\pi\|^2 \|S'\Sigma S - p'p/n\| + O_p(\epsilon_n^2) + O_p(1)\|\tilde{G}-G\|^2/n \\ &\leq [O_p(\epsilon_n^2) + O_p(1)\|\tilde{G}-G\|^2/n](1 + \|\Sigma - \bar{P}'\bar{P}/n\|) = O_p(\epsilon_n^2) + O_p(1)\|\tilde{G}-G\|^2/n. \quad \blacksquare\end{aligned}$$

The next results give convergence rates for sample average with dimension that can grow with sample size. For μ in Assumption 1 and s in Lemma 9.6 below, let $\Delta > 2\mu/(\mu-1)$ be as small as desired, and

$$\omega \equiv \begin{cases} -(1/2) + (1/s) - (1/\Delta), & s \leq 2\mu/(\mu-1) \\ -1/2, & s > 2\mu/(\mu-1) \end{cases} \quad \psi \equiv \begin{cases} 1 - (s/2\Delta), & s \leq 2\mu/(\mu-1) \\ 1/2, & s > 2\mu/(\mu-1) \end{cases}.$$

Lemma 9.6: If Assumption 3.1 is satisfied and there exists increasing $\nu_y(J)$ and B_{yi} with $\max_{j \leq J} |y_{ij}| \leq \nu_y(J) B_{yi}$, ($J = 1, 2, \dots$), and $|B_{yi}|_s < \infty$ for $s > 1$, then

$$\|\sum_1 y_i/n - E[y_i]\| = O_p(J^\psi \nu_y(J) n^\omega).$$

Proof: The proof for the case $s > 2\mu/(\mu-1)$ follows immediately from applying Davydov's inequality to the covariance terms in $E[\|\sum_1 y_i/n - E[y_i]\|^2]$. The proof for the other case follows by a truncation argument analogous to that used to prove weak laws of large numbers. ■

For the $P^K(q)$ of Assumption 3.4 and $\bar{K} = \bar{K}(n)$ of Assumption 3.5, let $\bar{P}(q) = P^{\bar{K}}(q)$, $\Sigma = E[\bar{P}(q_i)\bar{P}(q_i)']$, $\hat{\Sigma} = \sum_{t=1}^n \bar{P}(q_i)\bar{P}(q_i)'/n$, and $\tilde{K}$ denote the number of elements of $\bar{P}(q)\bar{P}(q)'$ that are nonzero at any point in Q .

Lemma 9.7: If Assumptions 3.1, 3.4, and 3.7 are satisfied then

$$\|\hat{\Sigma} - \Sigma\| = O_p(\tilde{K}\zeta_0(\bar{K})^4/n).$$

Proof: Apply Lemma 9.6 for $y_j = P_{k\bar{K}}(q)P_{\ell\bar{K}}(q)$ for all k and ℓ such that $P_{k\bar{K}}(q)P_{\ell\bar{K}}(q)$ is nonzero for some $q \in Q$. Here, $J = \tilde{K}$. Note that for all $q \in Q$, $|P_{k\bar{K}}(q_i)P_{\ell\bar{K}}(q_i)| \leq C\zeta_0(\bar{K})^2$, so that Assumption 9.1 is satisfied for $\nu_y(J) = \zeta_0(\bar{K})^2$, $B_{yi} = 1$, and $s = \infty$. Thus, the conclusion follows by Lemma 9.6 with $s = > 2\mu/(\mu-1)$. ■

Let y , $\hat{h}$, and h be $n \times J$ matrices with respective ij^{th} elements y_{ij} , $\hat{h}_j(q_i)$, and $h_j(q_i)$.

Lemma 9.8: If Assumptions 3.1, 3.4, 3.5, 3.7, and 9.1 are satisfied, then

$$(9.5) \quad \text{tr}[(y-h)'p(p'p)^{-1}p'(y-h)/n] = O_p(J\bar{K}\nu_y(J)^2/n).$$

Proof: Let $u \equiv y-h$, $P_i \equiv P^{\bar{K}}(q_i)$, $\Sigma \equiv E[P_i P_i']$, and $C_{jt} \equiv E[P_i P_{i+t}' u_{ij} u_{i+t,j}]$, ($0 \leq t \leq n-i$). By Assumption 9.1 and iterated expectations, $E[P_i P_i' u_{ij}^2] = E[P_i P_i' E[u_{ij}^2 | q_i]] \leq \nu_y(J)^2 \Sigma$. Also, $E[P_i u_{ij}] = 0$ by each element of P_i an element of $\mathcal{H}$ and orthogonality of the projection residual with $\mathcal{H}$. Then under the uniform mixing condition of Assumption 9.1 a), it follows by Lemma 2.2 of White and Domowitz (1984) that for any conformable ν and $c(t) \equiv Ct^{-\mu/2}$.

$$(9.6) \quad \nu' [C_{jt} + C_{jt}'] \nu \leq 2Ct^{-\mu/2} \nu' E[P_i P_i' u_{ij}^2] \nu \leq 2c(t) \nu_y(J)^2 \nu' \Sigma \nu,$$

Also, under Assumption 9.1 b), for $c(t)$ given there,

$$(9.7) \quad \nu' [C_{jt} + C_{jt}'] \nu \leq 2|E[\nu' P_i P_{i+t}' E[u_{ij} u_{i+t,j} | q_i, q_{i+t}]]| \leq 2\nu_y(J)^2 c(t) \nu' \Sigma \nu.$$

Thus, $2\nu_y(J)^2 c(t) \Sigma - (C_{jt} + C_{jt}')$ is positive semi-definite in either case, for $c(t)$ as in (9.6) or (9.7). Let $u_j = (u_{1j}, \dots, u_{nj})'$. Then it follows that

$$(9.8) \quad E[\|\Sigma^{-1/2} \tilde{P}' u\|^2 / n^2] = \sum_{j=1}^J E[u_j' P \Sigma^{-1} P' u_j] / n^2 \\ \leq J \nu_y(J)^2 \text{tr}(\Sigma^{-1/2} [C + 2 \sum_{t=0}^{\infty} c(t)] \Sigma \Sigma^{-1/2}) / n = O(J\bar{K}\nu_y(J)^2/n),$$

giving $\|\Sigma^{-1/2} P' u / n\| = O_p((J\bar{K})^{1/2} \nu_y(J) n^{-1/2})$. The conclusion then follows by Lemma 9.3, Lemma 9.7, Assumption 3.5, and Assumption 3.7. ■

Define

$$\delta(h, K, d, \nu) \equiv \inf_{\pi} \max_{|\lambda| \leq d} \{E[|\partial^\lambda \{h(q_i) - \pi' p^K(q_i)\}|^\nu]\}^{1/\nu} + \exp(-\exp(K)),$$

$$\delta(h, K, d, \infty) \equiv \inf_{\pi} \max_{|\lambda| \leq d} \sup_{q \in Q} |\partial^\lambda [h(q_i) - \pi' p^K(q)]| + \exp(-\exp(K)).$$

Lemma 9.9: If Assumptions 3.1, 3.4, 3.5, 3.7, and 9.1 are satisfied,

$$n^{-1/2} \|\hat{h} - h\| = O_p((J\bar{K})^{1/2} \nu_y(J) n^{-1/2} + \{\sum_j [\sum_{\underline{K} \leq K \leq \bar{K}} \delta(h_j, K, 0, v)^v]^{2/v}\}^{1/2}),$$

Proof: Let π_{jK} be such that $\{E[|h_j(q_i) - \pi_{jK}^K(q_i)|]^v\}^{1/v} \leq \delta(h_j, K, 0, v)$, and let $\pi_K = [\pi_{1K}, \dots, \pi_{JK}]$. By Assumption 3.5, $\delta(h_j, K, 0, v) \leq \delta(h_j, \underline{K}, 0, v)$, for $\underline{K} \leq K \leq \bar{K}$. Then for $1_{\hat{K}}$ the indicator function for the event $\underline{K} \leq \hat{K} \leq \bar{K}$,

$$\begin{aligned} E[1_{\hat{K}} \|h - p' \pi_{\hat{K}}\|^2 / n] &\leq E[\max_{\underline{K} \leq K \leq \bar{K}} \|h - p' \pi_K\|^2 / n] \\ &\leq \sum_j \{E[\sum_{\underline{K} \leq K \leq \bar{K}} |h_j(q_i) - \pi_{jK}^K(q_i)|^v]\}^{2/v} \leq \sum_j \{\sum_{\underline{K} \leq K \leq \bar{K}} \delta(h_j, K, 0, v)^v\}^{2/v}. \end{aligned}$$

Then by $\hat{1}_K = 1$ w.p.a.1 and the Markov inequality, $\|h - p' \pi_{\hat{K}}\|^2 / n \leq O_p(\sum_j \{\sum_{\underline{K} \leq K \leq \bar{K}} \delta(h_j, K, 0, v)^v\}^{2/v})$. Also, by Lemma 9.8, $(y - h)' p(p' p)^{-1} p' (y - h) / n = O_p((J\bar{K})^{1/2} \nu_y(J) n^{-1/2})$. The conclusion then follows by Lemma 9.4. ■

Lemma 9.10: If Assumptions 3.1, 3.4, 3.5, 3.7, and 9.1 are satisfied, then for the distribution $F(q)$ of q ,

$$\begin{aligned} &\{\sum_j \int [\hat{h}_j(q) - h_j(q)]^2 dF(q)\}^{1/2} \\ &= O_p((J\bar{K})^{1/2} \nu_y(J) n^{-1/2} + \{\sum_j [\sum_{\underline{K} \leq K \leq \bar{K}} \delta(h_j, K, 0, v)^v]^{2/v}\}^{1/2}). \end{aligned}$$

Proof: Let π_{jK} be as defined in the proof of the previous Lemma, with $P^K(q)$ replacing $p^K(q)$. Then by the same argument as there,

$$\int 1_{\hat{K}} \|h(q) - P^{\hat{K}}(q)' \pi_{\hat{K}}\|^2 dF(q) \leq \sum_j \{\sum_{\underline{K} \leq K \leq \bar{K}} \delta(h_j, K, 0, v)^v\}^{2/v}.$$

Next, apply Lemma 9.5 with $\Sigma \equiv \int P^{\bar{K}}(q) P^{\bar{K}}(q)' dF(q)$, $\bar{P} = [P^{\bar{K}}(q_1), \dots, P^{\bar{K}}(q_n)]'$, $P = [P^{\hat{K}}(q_1), \dots, P^{\hat{K}}(q_n)]'$, S the selection matrix such that $P^{\hat{K}}(q) = S' P^{\bar{K}}(q)$ w.p.a.1, $\hat{\pi} = (P' P)^{-1} P' y$, $\pi = \pi_{\hat{K}}$, $G = h$, $\epsilon_n = (J\bar{K})^{1/2} \nu_y(J) n^{-1/2}$. From the

conclusion of Lemma 9.5 and the argument in the previous Lemma, it follows that

$$\begin{aligned}
\int \|P^{\hat{K}}(q)[\hat{\pi}-\pi_{\hat{K}}]\|^2 dF(q) &= \int \|P^{\hat{K}}(q)[\hat{\pi}-\pi_{\hat{K}}]\|^2 dF(q) \\
&= \text{tr}\{(\hat{\pi}-\pi_{\hat{K}})' [\int P^{\hat{K}}(q)P^{\hat{K}}(q)' dF(q)] (\hat{\pi}-\pi_{\hat{K}})\} \\
&= \text{tr}[(\hat{\pi}-\pi)' S' \Sigma S (\hat{\pi}-\pi)] \leq O_p(\epsilon_n^2) + O_p(1) \|G-P\pi\|^2/n \\
&= O_p(\epsilon_n^2) + O_p(1) \|h-P'\pi_{\hat{K}}\|^2/n = O_p(\epsilon_n^2) + O_p(\sum_j \{\sum_{\underline{K} \leq K \leq \bar{K}} \delta(h_j, K, 0, v)^v\}^{2/v}).
\end{aligned}$$

Then by the first equation of this proof, and $1_{\hat{K}} = 1$ w.p.a.1,

$$\int \|\hat{h}(q)-h(q)\|^2 dF(q) \leq C\{\int \|P^{\hat{K}}(q)[\hat{\pi}-\pi_{\hat{K}}]\|^2 dF(q) + \int 1_{\hat{K}} \|h(q)-P^{\hat{K}}(q)'\pi_{\hat{K}}\|^2 dF(q)\},$$

so that the conclusion follows. ■

Lemma 9.11: If Assumptions 3.1, 3.4, 3.5, 3.7, and 9.1 are satisfied, $h(q)$ and $P_{kK}(q)$ are differentiable of order $|\lambda|$ for each $k \leq K$ and K , $\max_{k \leq K} \sup_{q \in Q} |\partial^\lambda P_{kK}(q)| \leq \zeta_{|\lambda|}(K)$, and $\zeta_{|\lambda|}(K) \geq \zeta_0(K)$, then

$$\sup_{q \in Q} \|\partial^d \hat{h}(q) - \partial^d h(q)\| = O_p(\bar{K}^{1/2} \zeta_{|\lambda|}(\bar{K}) [(\bar{K}/n)^{1/2} + \{\sum_j \delta(h_j, \underline{K}, |\lambda|, \infty)^2\}^{1/2}]).$$

Proof: Let $\pi_{jK} \in \mathbb{R}^K$ be such that, for $\Delta_{jK}(q) \equiv h_j(q) - P^K(q)'\pi_{jK}$, $\sup_{q \in Q} |\Delta_{jK}(q)| \leq \delta(h_j, K, |\lambda|, \infty)$ and $\sup_{q \in Q} |\partial^\lambda \Delta_{jK}(q)| \leq \delta(h_j, K, |\lambda|, \infty)$ for each j and K . Also, let π_K be the $K \times J$ matrix with j^{th} column π_{jK} and Δ_K the $n \times J$ matrix with ij^{th} element $\Delta_{jK}(q_i)$. Note that by Assumption 3.7, $\delta(h_j, \hat{K}, |\lambda|, \infty) \leq \delta(h_j, \underline{K}, |\lambda|, \infty)$ w.p.a.1. Thus, for $P \equiv [P^{\hat{K}}(q_1), \dots, P^{\hat{K}}(q_n)]'$, w.p.a.1.

$$(9.9) \quad \|\hat{h}-P\pi_{\hat{K}}\|^2/n = \|\hat{\Delta}_{\hat{K}}\|^2/n = \sum_j \sum_i \Delta_{j\hat{K}}(\hat{q}_i)^2/n \leq \sum_j \delta(h_j, \underline{K}, |\lambda|, \infty)^2.$$

Next, let $Y = y$ and $G = h$. Note that columns of P are a nonsingular linear transformation of the columns of p , so that $P(P'P)^{-1}P' = p(p'p)^{-1}p'$. Thus, by Lemma 9.8, $\text{tr}[(Y-G)'P(P'P)^{-1}P'(Y-G)]/n = O_p(\bar{K}/n)$. Thus, by Lemma 9.5,

$$(9.10) \quad \|\hat{\pi} - \pi_{\hat{K}}\|^2 \leq O_p(\bar{K}/n) + O_p(1) \sum_j \delta(h_j, \underline{K}, |\lambda|, \infty)^2 \\ = O_p([\bar{K}/n]^{1/2} + \{\sum_j \delta(h_j, \underline{K}, |\lambda|, \infty)^2\}^{1/2})^2.$$

Noting that $\partial^{\lambda \hat{h}}(q) = \hat{\pi}' \partial^{\lambda \hat{K}}(q)$, it then follows by the Cauchy-Schwartz inequality that for any $q \in Q$,

$$\|\partial^{\lambda \hat{h}}(q) - \partial^{\lambda h}(q)\|^2 \leq C\{\|(\hat{\pi} - \pi_{\hat{K}})' \partial^{\lambda \hat{K}}(q)\|^2 + \|\pi_{\hat{K}}' \partial^{\lambda \hat{K}}(q) - \partial^{\lambda h}(q)\|^2\} \\ \leq \|\hat{\pi} - \pi_{\hat{K}}\|^2 \|\partial^{\lambda \bar{K}}(q)\|^2 + \sum_j \Delta_{j\hat{K}}(q)^2 \leq \bar{K} \zeta_{|\lambda|}(\bar{K})^2 \|\hat{\pi} - \pi_{\hat{K}}\|^2 + \sum_j \delta(h_j, \underline{K}, |\lambda|, \infty).$$

Since the term following the last inequality does not depend on n , the first conclusion then follows from eq. (9.12). ■

References

- Abramowitz, M. and Stegun, I. A., eds. (1972). *Handbook of Mathematical Functions*. Washington, D.C.: Commerce Department.
- Andrews, D.W.K. (1991). Asymptotic normality of series estimators for various nonparametric and semiparametric models. *Econometrica*. 59 307-345.
- Andrews, D.W.K. and Whang, Y.J. (1990). Additive interactive regression models: Circumvention of the curse of dimensionality. *Econometric Theory*. 6 466-479.
- Bickel P., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1990): "Efficient and Adaptive Inference in Semiparametric Models" monograph, forthcoming.
- Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*. 80 580-598.
- Breiman, L., Stone, C.J. (1978). Nonlinear additive regression. note.
- Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *Annals of Statistics*. 17 453-510.
- Burman, P. and Chen, K.W. (1989). Nonparametric estimation of a regression function. *Annals of Statistics*. 17 1567-1596.
- Chamberlain, G. (1986). Notes on semiparametric regression. Preprint. Department of Economics, Harvard University.
- Davydov, Y.A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and Its Applications*. 13 691-696.
- Engle, R.F., Granger, C.W.J., Rice, J., and Weiss, A. (1984). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*. 81 310-320.
- Friedman, J. and Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*. 76 817-823.
- Gallant, A.R. (1980). Explicit estimators of parametric functions in nonlinear regression. *Journal of the American Statistical Association*. 75 182-193.
- Gallant, A.R. (1981). On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form. *Journal of Econometrics*. 76 211 - 245.
- Hansen, L.P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators, *Journal of Econometrics*. 30 203-238.

- Hardle, W. and Stoker, T. (1989). Investigation of smooth regression by the method of average derivatives. *Journal of the American Statistical Associations.* 84 986-995.
- Heckman, N.E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society, Series B.* 48 244-248.
- Lorentz, G.G. (1986). *Approximation of Functions*. New York: Chelsea Publishing Company.
- Newey, W.K. (1988a). Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics.* 38 301-339.
- Newey, W.K. (1988b). Two-step series estimation of sample selection models. preprint. Princeton University. Department of Economics.
- Newey, W.K. (1990). Series estimation of regression functionals. preprint. MIT Department of Economics.
- Newey, W.K. (1991). The asymptotic distribution of semiparametric estimators. Preprint. MIT. Department of Economics.
- Powell, M.J.D. (1981). *Approximation Theory and Methods*. Cambridge, England: Cambridge University Press.
- Powell, J.L., Stock, J.H., and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica.* 57 1403-1430.
- Rice, J. (1986). Convergence rates for partially splined estimates. *Statistics and Probability Letters.* 4 203-208.
- Robinson, P. (1988). Root-n-consistent semiparametric regression. *Econometrica.* 56 931-954.
- Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *Annals of Statistics.* 14 1139-1151.
- Schumaker, L.L. (1981): *Spline Functions: Basic Theory*. Wiley, New York.
- Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics.* 10 1040-1053.
- Stone, C.J. (1985). Additive regression and other nonparametric models. *Annals of Statistics.* 13 689-705.
- Stone, C.J. (1990). L_2 rate of convergence for interaction spline regression," Tech. Rep. No. 268, Berkeley).
- Wahba, G. (1984). Cross-validated spline methods for the estimation of multivariate functions from data on functionals. In *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory* (H.A. David and H.T. David, eds.) 205-235, Iowa State University Press, Ames, Iowa.

- White, H. (1980). Using least squares to approximate unknown regression functions. *International Economic Review*. 21 149-170.
- White, H. (1984). *Asymptotic Theory for Econometricians*. Orlando: Academic Press.
- White, H. and Domowitz, I. (1984). Nonlinear regression with dependent observations. *Econometrica*. 52 143-161.
- Zeldin, M.D. and Thomas, D.M. (1975). Ozone trends in the Eastern Los Angeles basin corrected for meteorological variations. *Proceedings International Conference on Environmental Sensing and Assessment*, 2, held September 14-19, 1975, in Las Vegas, Nevada.



Date Due

MAY 31 1997

DEC 31 1999

Lib-26-67

MIT LIBRARIES



3 9080 00719551 1

